

STIC-Biotech/ChemLib

174960

From: Ramirez, Delia  
Sent: Tuesday, December 20, 2005 4:43 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/049,750

Hi,

I would like to request the following search:

1. a standard search of SEQ ID NO: 11, 13 and 14 in the protein databases (commercial & interference)

Thank you very much,

Delia M. Ramirez, Ph.D.  
Patent Examiner  
Recombinant Enzymes-Art Unit 1652  
USPTO  
400 Dulany Street, Remsen Bldg., 2D74, Mail room 2C70  
Alexandria, VA 22314  
(571) 272-0938  
delia.ramirez@uspto.gov

RECEIVED  
DEC 20 2005  
STIC/GEN. DIVISION  
(STIC)

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:05:55 ; Search time 172.852 Seconds  
(without alignments)  
4682.260 Million cell updates/sec

Title: US-10-049-750-11

Perfect score: 1590

Sequence: 1 atggatgcgcgggtattgg.....atcttcgtgaacgaataa 921

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlp  
-Q/cn2\_1/USPTO spool\_p/US10049750/runat\_23122005\_113551\_6285/app\_query.fasta\_1.1742  
-DB=A\_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10049750 @CN 1 1 476 @runat\_23122005\_113551\_6285 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_21:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*  
9: Geneseq2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1570	98.7	306	4 AAB67588	Aab67588 Amino aci
2	1519	95.5	306	4 AAB52865	Abd52865 Escherich
3	1207	75.9	311	7 ADF05492	Adf05492 Bacterial
4	870	54.7	298	8 ADS25796	Ads25796 Bacterial
5	870	54.7	298	8 ADS22364	Ads22364 Bacterial
6	870	54.7	298	8 ADS25248	Ads25248 Bacterial
7	870	54.7	298	8 ADS25978	Ads25978 Bacterial
8	831.5	52.3	321	4 AAU60867	Aau60867 Propionib
9	831.5	52.3	321	6 ABM57386	Abm57386 Propionib

10	797	50.1	304	8 ADN25077	Adn25077 Bacterial
11	550	34.6	308	4 AAU36261	Aau36261 Pseudomon
12	549	34.5	503	7 ABO83136	Abob3136 Pseudomon
13	546	34.3	322	4 AAB73533	Aab73533 Human tra
14	546	34.3	322	5 ABP55421	Abp55421 Human pro
15	546	34.3	322	4 ADO89874	Ado89874 Antagonis
16	546	34.3	322	8 ADR05197	Adr05197 Human rib
17	546	34.3	322	9 ADY17572	Ady17572 PRO polyp
18	546	34.1	322	4 AAM38771	Aam38771 Human pol
19	523	32.9	343	8 ADN22931	Adn22931 Bacterial
20	507	31.9	309	8 ADS23317	Ads23317 Bacterial
21	506	31.8	302	8 ADN24730	Adn24730 Bacterial
22	506	31.8	303	8 ADN21972	Adn21972 Bacterial
23	501	31.5	299	5 ABB90150	Abb90150 Human pol
24	499.5	31.4	300	8 ADN24942	Adn24942 Bacterial
25	499.5	31.4	303	8 ADN22183	Adn22183 Bacterial
26	495.5	31.2	300	8 ADS24911	Ads24911 Bacterial
27	491.5	30.9	320	7 ADF05698	Adf05698 Bacterial
28	488.5	30.7	293	8 ADN26498	Adn26498 Bacterial
29	476.5	30.0	304	4 AAU37424	Aau37424 Staphyloc
30	476.5	30.0	304	4 AAU37560	Aau37560 Staphyloc
31	474.5	29.8	304	6 ABM71981	Abm71981 Staphyloc
32	474.5	29.8	293	8 ADS44895	Ads44895 Salmonell
33	468.5	29.5	309	4 AAU38290	Aau38290 Salmonell
34	468.5	29.5	318	8 ADS44168	Ads44168 Bacterial
35	467	29.4	309	8 ADN17558	Adn17558 Bacterial
36	460	28.9	294	8 ADS28417	Ads28417 Bacterial
37	460	28.9	304	4 ABB68082	Abb68082 Drosophil
38	458	28.8	302	4 AAG82984	Aag82984 S. epider
39	457.5	28.8	293	8 ADS27571	Ads27571 Bacterial
40	454	28.6	303	8 ADS41653	Ads41653 Bacterial
41	452.5	28.5	330	7 ABO67526	Abob67526 Klebsiell
42	452	28.4	307	6 ABP19021	Abp19021 Pathogen
43	452	28.4	315	5 ABP39676	Abp39676 Staphyloc
44	452	28.4	315	8 ADS05785	Ads05785 Staphyloc
45	451	28.4	271	4 AAM40557	Aam40557 Human pol

## ALIGNMENTS

RESULT 1

AAB67588

ID AAB67588 standard; protein; 306 AA.

XX

AC AAB67588;

XX

DT 29-MAY-2001 (first entry)

XX

DE Amino acid sequence of a deoxyribokinase enzyme.

XX

KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;  
KW purine nucleoside phosphorylase; phosphopentose mutase;  
KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;  
KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.

XX

OS Salmonella typhi.

XX

PN WO200114566-A2.

XX

PD 01-MAR-2001.

XX

PF 18-AUG-2000; 2000WO-EP008088.

XX

PR 20-AUG-1999; 99EP-00116425.

XX

PA (HOF) ROCHE DIAGNOSTICS GMBH.

PA (INSP) INST PASTEUR.

PA (PHAR-) PHARMA-WALDHOF GMBH & CO KG.

XX

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;

PI Marliere P, Pochet S;

XX WPI; 2001-235026/24.

DR

DR N-PSDB; AAF55444.  
 XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting  
 PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside  
 PT and an inorganic phosphate.  
 XX  
 PS Disclosure; Page 59-61; 73pp; English.  
 XX  
 CC The present sequence represents a deoxyribokinase enzyme. This enzyme is  
 CC involved in the biosynthesis of deoxyribonucleosides, and is used in the  
 CC method of the invention. The specification describes a method for the in  
 CC vitro enzymatic synthesis of deoxyribonucleosides. The method comprises  
 CC reacting deoxyribose 1-phosphate and a nucleobase to form a  
 CC deoxyribonucleoside and an inorganic phosphate. Enzymes which may be used  
 CC in the method of the invention include thymidine phosphorylase, purine  
 CC nucleoside phosphorylase, phosphopentose mutase, phosphopentose aldolase,  
 CC fructose 1,6-diphosphate aldolase, deoxyribokinase, and nucleoside 2-  
 CC deoxyribosyltransferase  
 XX  
 SQ Sequence 306 AA;

Alignment Scores:  
 Pred. No.: 1.14e-168 Length: 306  
 Score: 1570.00 Matches: 306  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 98.74% Indels: 0  
 DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x AAB67588 (1-306)

QY 1 ATGGATATCGCGGTATTGGCTTACATGGTGGACCTTATCACCTACACCAACGATG 60  
 DB 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrThrAsnGlnMet 20  
 QY 61 CCCAAGAAGGGGAACTCTGGAAGCGCGGGCTTTAAATCGGCTCGCGCGGAAAGGG 120  
 DB 21 ProlysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyLysGly 40  
 QY 121 GCGAACACGCGGTGGCGCGCTAAGCTCAATCAAAAGTATTGATGTGACCAAGTG 180  
 DB 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60  
 QY 181 GCGACGATATTTTCCGACACACCATTCGATCTCGAATCTCGGGGGATCAATCG 240  
 DB 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80  
 QY 241 ACCTATGTAGAAAAGTACCGTGTACAGCAGCGCGGTAGCGCGGATTTTCGTCAACGCC 300  
 DB 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100  
 QY 301 AACTCCAGCAACAGCATTCGATCATCAAGGCGCTAAACAAGTTCTCTCGCGGGAAGAT 360  
 DB 101 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120  
 QY 361 ATCGATCGCGCGGCGAAGATTTAAATAATGCCAGCTTATTCGTCGCACTGGAAGTT 420  
 DB 121 IleAspArgAlaAlaGluAspLeuLysCysGlnLeuIleValLeuGlnLeuGluVal 140  
 QY 421 CAGCTTGAACCGTTTATCAGCAATAGATTTTGGCAAGAAACACCGGATTTGAAGTGTTA 480  
 DB 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysHisGlyIleGluValLeu 160  
 QY 481 TTAACCCCTCGCGCACATACGGGAATTAGATATGCTCTTATGCTGCTGTAATGCGATTTC 540  
 DB 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180  
 QY 541 TTTGTACCTAATCAACCGCTGGAATATTAACCGGTATGCCAGTGGATACCATATGAC 600  
 DB 181 PheValProAsnGluThrGluLeuGluIleLeuThrGlyWecProValAspThrTyrAsp 200  
 QY 601 CATATTTCGCGCAGCGCGCTTCGCTGGTAGATAAAGGGCTGAACAATATTATTGTCACC 660

DB 201 HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220  
 QY 661 ATGGCGAGAAAGCGCGCTGTGGATGACGCTGACAGGAGTCCATGTTCCGCGCGTTT 720  
 DB 221 MetGlyGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240  
 QY 721 AGAGTGAACCGCTGTGTATACCAAGCGCGCGCGATGCTTTATCGGCTGTTTCGCGCAT 780  
 DB 241 ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260  
 QY 781 TACTACGTCCAGAGCGGGGATGTGAAGCGGCATGAAAGCCGCTCTCTTTCGCGCT 840  
 DB 261 TyrTyrValGlnSerGlyAspValGluAlaAlaMetLysLysAlaValLeuPheAlaA 280  
 QY 841 TTCACGCTCACCGGAAAGGACCAATCTCTTATCCAAGCATTGAGCAATTTAATGAG 900  
 DB 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300  
 QY 901 TATCTTTCTGTTGAACGAA 918  
 DB 301 TyrLeuSerLeuAsnGlu 306

RESULT 2

ABBS2865  
 ID ABBS2865 standard; protein; 306 AA.

XX ABBS2865;

DT 11-FEB-2002 (first entry)

DE Escherichia coli polypeptide SEQ ID NO 1139.

KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;  
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;  
 KW systemic infection; non-diarrhoeal infection; septicemia;  
 KW pyelonephritis; antibiotic resistance.

OS Escherichia coli.

PN WO200166572-A2.

PD 13-SEP-2001.

PF 12-MAR-2001; 2001WO-EP003445.

PR 10-MAR-2000; 2000FR-00003145.

PR 02-FEB-2001; 2001FR-00001449.

(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX WPI; 2001-550253/61.

XX A library of DNA fragments of Escherichia coli strains for the phylogenetic  
 PT determination of a given strain comprises polynucleotides of nature B2/D+  
 XX A-.

PS Example 6; Fig 6; 646pp; English.

XX The invention relates to a library of DNA fragments of Escherichia coli  
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and  
 CC encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature  
 CC B2/D+A-. The polynucleotides have potential antiinflammatory.  
 CC antibacterial and immunosuppressive activity as part of pharmaceutical  
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli  
 CC infections. The polypeptides are useful for determining the phylogenetic  
 CC group of a given E. coli strain. These polypeptides can detect and treat  
 CC an undesired development of E. coli, particularly an extra-intestinal  
 CC infection that include systemic and non-diarrhoeal infections such as  
 CC septicemia, pyelonephritis and meningitis this is particularly  
 CC advantageous as bacterial resistance is increasing with the more frequent  
 CC use of broad spectrum antibiotics



```
XX SQ Sequence 306 AA;
Alignment Scores:
Pred. No.: 7,13e-163 Length: 306
Score: 1519.00 Matches: 294
Percent Similarity: 98.69% Conservatives: 8
Best Local Similarity: 96.08% Mismatches: 4
Query Match: 95.53% Indels: 0
DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x ABB52865 (1-306)
QY 1 ATGGATATCGCGGTTATTTGGCTTAACATGGTGGACCTTATCACCTACCAACCAAGATG 60
Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrThrAsnGlnMet 20
QY 61 CCCAAGAGAGGGAACCTCTGGAAGCGCGCGGTAAAAATCGGCTGCGCGGAAGGG 120
Db 21 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyLysGly 40
QY 121 GCGAACCGCGCGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGG 180
Db 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
QY 181 GCGACGATATTTTCCGCGACACACCATTCGTAATCTCGAATCTCTGGGGGATCAATACG 240
Db 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
QY 241 ACGTATGTAGAAAAGTACGCTGACAGAGCGCGGTAGCGCGATTTTCGTCAACGCC 300
Db 81 ThrTyValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
QY 301 AACTCCAGCAACAGCATTCTGATCATCAAGGCGCTAACAAAGTTTCTCTCGCGGAAGAT 360
Db 101 AsnSerAsnSerIleLeuIleIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
QY 361 ATCGATCGCGCGCGGAAGATTTAAAAAATGCGAGCTTATTGTTCTGCAACTGGAAGTT 420
Db 121 IleAspArgAlaAlaGluAspLeuLysLysCysLeuLysIleValLeuGlnLeuVal 140
QY 421 CAGCTTGAACGGTTTATCAGCAATAGAAATTTGGCAAGAACACCGGATTTGAAGTTA 480
Db 141 GlnLeuGluThrValTyHisAlaIleGluPheGlyLysLysLysIleGluValLeu 160
QY 481 TTAACCCCTGCGCCAGCATTACGGAAATAGATATCTTATGCTGTAATGCGGATTC 540
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyAlaCysLysCysAspPhe 180
QY 541 TTTGTACCTTAATGAAACCGAGCTGGAATATTAAACCGGTATGCGATGATACCTATGAC 600
Db 181 PheIleProAsnGluThrGluLeuGluIleLeuThrGlyMetSerValAspThrTyAsp 200
QY 601 CATATTCGCGCAGCGCAGCTTCGCTGGTAGATAAAGGGCTGAACAAATATTATGTCACC 660
Db 201 HisIleArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleValThr 220
QY 661 ATGGGCGAGAAAGCGCGCTGTGGATGACCGCTGACAGGAAGTCCATGTCGCGGTTT 720
Db 221 MetSerGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
QY 721 AGAGTCAACGCTTGTATACACCGCGCGCGGCGATGCCCTTTATCGGCTGTTTCGCGCAT 780
Db 241 LysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 260
QY 781 TACTACGTCCAGACCGGGATGGAAGCGGCATGAAAGCCGCTCTCTTTGCGCGT 840
Db 261 TyTyValGlnSerGlyAspValGluAlaAlaLeuLysLysAlaAlaLeuPheAlaAla 280
QY 841 TTCAGCGTACCAGGAAAGCCCAATCTCTTATCCAGCATTCAGCAATTTATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyProSerIleGluGlnPheAsnGlu 300

901 TATCTTTTCGTTCAACGAA 918
Db 301 PheLeuThrLeuAsnGlu 306

RESULT 3
ADF05492
ID ADF05492 standard; protein; 311 AA.
XX AC ADF05492;
XX DT 12-FEB-2004 (first entry)
XX DE Bacterial polypeptide #1605.
XX KW Proteus mirabilis infection; bacterial infection; antibacterial;
XX OS Proteus mirabilis.
XX PN US6605709-B1.
XX PD 12-AUG-2003.
XX PF 05-APR-2000; 2000US-00543681.
XX PR 09-APR-1999; 99US-0128706P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL;
XX DR WPI; 2003-895291/82.
XX DR N-PSDB; ADF01320.
XX PT New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX PS Disclosure; SEQ ID NO 5777; 870pp; English.
XX CC The invention relates to new Proteus mirabilis polypeptides and
CC polynucleotides. The invention also relates to antibodies against the
CC polypeptides, methods for producing the polypeptides, a method of
CC generating vaccines for immunising an individual against P. mirabilis, a
CC method for evaluating a compound for the ability to bind a P. mirabilis
CC polypeptide and a method for screening test compounds for anti-bacterial
CC activity. The polypeptides and polynucleotides are useful as molecular
CC targets for diagnosing, preventing and treating pathological conditions
CC resulting from bacterial infection, as reagents for diagnosis of
CC bacterial diseases, as components of antibacterial vaccines, as targets
CC for antibacterial drugs or as bio-control agents for plants. This
CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX SQ Sequence 311 AA;

Alignment Scores:
Pred. No.: 2,05e-127 Length: 311
Score: 1207.00 Matches: 232
Percent Similarity: 87.79% Conservatives: 34
Best Local Similarity: 76.57% Mismatches: 37
Query Match: 75.91% Indels: 0
DB: 7 Gaps: 0

US-10-049-750-11 (1-921) x ADF05492 (1-311)
QY 1 ATGGATATCGCGGTTATTTGGCTTAACATGGTGGACCTTATCACCTACCAACCAAGATG 60
Db 6 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrThrTyIleAspArgMet 25
QY 61 CCCAAGAGAGGGAACCTCTGGAAGCGCGCGGTAAAAATCGGCTGCGCGGAAGGG 120
Db 26 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyLysGly 45
```

QY 121 GCGAACCGCGCTGCGCGCGCTAAGCTCAATTCAAAGATATTGATGTTGACCAAGTG 180  
Db 46 AlaAenGlnAlaValAlaAlaAlaLysLeuAenSerLysValMetMetLeuThrLysVal 65  
QY 181 GCGACGATATTTTGGCCGACACACCATTCGTAATCTCGAATCTCTGGGGATCAATACG 240  
Db 66 GlyAspAspIlePheAlaAspAnThrIleMetAenLeuGluSerTyrGlyIleAnThr 85  
QY 241 AGCTATGTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGCATTTTCGTCAACGCC 300  
Db 86 ArgTyrValGluLysValProGlyThrThrSerGlyValAlaProIlePheValThrSer 105  
QY 301 AACTCCAGCAACAGCATCTCTGATCATCAAGAGCGCTAACAACTTCTCTCGCGGAAGAT 360  
Db 106 GlnSerSerAsnSerIleLeuIleValLysGlyAlaAenGlnHisLeuSerAlaGluAsp 125  
QY 361 ATCGATCGCGCGCGGAAGATTAAAAAAATCCAGCTTATTGTTCTGCAACTGGAAGTT 420  
Db 126 IleAspArgAlaAlaValSerLeuLysLysCysArgMetIleValLeuGlnLeuGluIle 145  
QY 421 CAGCTTGAACCGTTTATCAGCAATAGAAATTTGGCAAGAAACACACGGGATTTGAAGTTA 480  
Db 146 ProLeuGluThrValTyrHisAlaIleGluPheGlyAsnLysHisIleProValLeu 165  
QY 481 TTTAAACCTCGCGCGCATTCACGGAATTAGATATGCTTATGCTCTGTAATGCGATTTC 540  
Db 166 PheAenProAlaProAlaSerLysAlaLeuAspLeuThrIleAlaAlaGlnCysAspPhe 185  
QY 541 TTTGTACCTTAATGAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600  
Db 186 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProIleAsnThrLeuAsp 205  
QY 601 CATATTCGCGCAGCGCATTCGCTGTGTAGATAAAGGCTGAAACAATATTATGTCACC 660  
Db 206 GluIleArgGluAlaAlaTyrPheLeuLeuGluLysGlyPheLysAsnIleIleValThr 225  
QY 661 ATGCGCGAGAAAGCGCGCTGTGGATGACGCGTACCAGGAAGTCCATGTTCCGCGGTTT 720  
Db 226 LeuGlyGluLysGlyAlaLeuIleValAenGlyGluIleGluLysTyrIleProAlaIle 245  
QY 721 AGAGTGAACGCTGTTGATACACGCGCGCGCGATGCTTTATCGGCTGTTTTCGCGCAT 780  
Db 246 GluValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 265  
QY 781 TACTAGCTCAGAGCGGGATGTGGAAGCGGCATGAAAAAAGCGCTCTCTTTCGCGCT 840  
Db 266 TyrTyrValHisThrGlyAsnIleGluGluAlaLeuAsnLysAlaValMetPheSerGly 285  
QY 841 TTCAGCGTCACCGGAAAGCGCACTCTTATCCAGCATTCGAGCAATTTAATGAG 900  
Db 286 LeuSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGluPheSerGlu 305  
QY 901 TATCTTTTCG 909  
Db 306 PheLeuSer 308

RESULT 4  
ADS25796  
ID ADS25796 standard; protein; 298 AA.  
XX  
AC ADS25796;  
XX

DT 02-DEC-2004 (first entry)

DE Bacterial polypeptide #14829.

XX Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

KW bacterial polypeptide.

XX Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide

XX for expression of a polynucleotide encoding a polypeptide from a

XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 14829; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polypeptide used in the  
CC scope of the invention. Note: The sequence data for this patent did not  
CC form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).

XX Sequence 298 AA;

Alignment Scores:  
Pred. No.: 3 99e-89 Length: 298  
Score: 870.00 Matches: 167  
Percent Similarity: 76.61% Conservative: 59  
Best Local Similarity: 56.61% Mismatches: 69  
Query Match: 54.72% Indels: 0  
DB: 8 Gaps: 0

US-10-049-750-11 (1-921) x ADS25796 (1-298)

QY 7 ATFCGCGTTATGGCTCTAACAATGGTGGACCTTATCACCACCAACAGATGCCAAA 66

Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 23

QY 67 GAAGGGGAACTCTGGAACCGCGCGGTTTAAATCGGCTCGCGGAAAGGGCGGAC 126

Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyLysGlyAlaAsn 43



```
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
QY 247 GTAGAAAAGTACCGGTACCGAGCGGTAGCGCGATTTTCGTCAACGCCACTCC 306
Db 84 ValValLysValSerGlyLysSerGlyValAlaProIlePheValGluGlnSerGly 103
QY 307 AGCAACAGCATTCGTATCATCAAGCGCTAACAGTTTCTCTCGCGCGAAGATATCGAT 366
Db 104 GluAsnSerIleLeuIleValLysGlyAlaAsnAlaAspLeuLeuProValGluValAsp 123
QY 367 CCGCGCGCGAAGATTAAAAATCCAGCTTATTGTTCTGCAACTGGAAGTTGAGTTT 426
Db 124 LysAlaAlaAspLeuLysGlyCysGlyLeuLeuMetGlnMetGluValProVal 143
QY 427 GAAACGGTTTATCAGCATATGAAATTTGCCAGAAACACGCGATTCGAAGTTTATAAC 486
Db 144 GluThrValThrHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
QY 487 CTGCGCGCAGCATTCAGCGAATTAGATATGCTTATGCTGTAATGCGATTTCTTGTGA 546
Db 164 ProAlaProAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
QY 547 CTTAATGAACCGAGCTGGAAATATTACCGGTATGCCAGTGGATACCTATGACCATATT 606
Db 184 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
QY 607 CCGCGACGCGCGCTGCGTGTAGATAAGGCTGAACAATATTATTCACCATGGGC 666
Db 204 ValArgAlaAlaArgSerLeuIleAlaArgGlyIleArgThrValIleValThrLeuGly 223
QY 667 GAGAAAGCGCGCTGGGATGACGCTGACAGGAAGTCCATGTTCCGCGCTTGAAGTG 726
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
QY 727 AACGCTGTTGATACCAACGCGCGCGCGATGCTTTATCGCTGTTTCCGCAATTACTAC 786
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
QY 787 GTTCAGAGCGGGATGTGAAGCGCGCATGAAAGCGCTCTCTTTCGCGCTTTCAGC 846
Db 264 AlaGluThrArgGluValValSerSerLeuLysLysAlaSerLeuTyrAlaAlaHisSer 283
QY 847 GTCACCGGGAAGGACCAACCTCTTATCCAAGCATTGAGCAA 891
Db 284 IleThrArgProGlyThrGlnLysAlaIlyAlaSerIleAspGlu 298

RESULT 6
ADS25248
ID ADS25248 standard; protein; 298 AA.
XX
AC ADS25248;
XX
DT
DE
DE Bacterial polypeptide #14281.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
XX 18-DEC-2003.
PD
PF 20-FEB-2003; 2003US-00369493.
XX
```

```
PR 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
DR
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 14281; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polypeptide used in the
XX scope of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 298 AA;
XX
Alignment Scores:
Pred. No.: 3,998-89 Length: 298
Score: 870.00 Matches: 167
Percent Similarity: 76.61% Conservative: 59
Best Local Similarity: 56.61% Mismatches: 69
Query Match: 54.72% Indels: 0
DB: Gaps: 0
XX
US-10-049-750-11 (1-921) x ADS25248 (1-298)
QY 7 ATCGCGGTTATTGGCTTAACATGGTGACCTTATCACCTACCAACAGATGCCCAA 66
Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 23
QY 67 GAAGGGGAAACTCTGGAAGCGCGCGTTTAAATTCGGCTCGCGGAAAGGGGGAAC 126
Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyClyLysGlyAlaAsn 43
QY 127 CAGGCGGCGCGCGCTTAAGCTCAATCAAAAGTATTGATTTGACCAAGTGGGCGAC 186
Db 44 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrValGlyAsp 63
QY 187 GATATTTTTCGCGACCAACACCATTCGTATCTCGAATCTCGGGGATCAATACGAGTAT 246
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
QY 247 GTAGAAAAGTACCGGTACCGAGCGGTAGCGCGATTTTCGTCAACGCCACTCC 306
XX
```

Db 84 ValValLysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103  
QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAAAGATTCTCTCGCGGGAAGATATCGAT 366  
Db 104 GluAenSerIleLeuIleValLysGlyAlaAenAlaAspLeuLeuProValGluValAsp 123  
QY 367 CGCGCGCGGAGATTAAATAATCCAGCTTATTGTTCTGCACTGGAAGTTCAGTT 426  
Db 124 LysAlaAlaAspLeuLysGlyLysGlyLeuLeuMetGlnMetGluValProVal 143  
QY 427 GAAACGGTTTATCACCAATAGATTGGCAAGAACCGGATTGAAGTGTATTAAAC 486  
Db 144 GluThrValTyHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAen 163  
QY 487 CTGCGCCAGCATTCAGGAATTAGATATGCTTATGCTGCTGTAATGCGATTCTTTGTA 546  
Db 164 ProAlaProAlaAlaAlaAenLeuAspProGluArgIleArgGlnValThrPheLeuVal 183  
QY 547 CCTATGAACCGAGCTGGAATATTAAACGGTATGCCAGTGATGATACCTATGACCATATT 606  
Db 184 ProAenGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203  
QY 607 CGCGCAGCGCGATTCGCTGCTAGATTAAGGCTGAACAATATTATTCACCATGGGC 666  
Db 204 ValArgAlaAlaArgSerIleIleAlaArgGlyIleArgThrValIleValThrLeuGly 223  
QY 667 GAGAAGGCGCGTGTGGATGACGCGTACAGGAAGTCCATGTTCCGCGTTTGAAGTG 726  
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAlaAsnIleGluProValLysVal 243  
QY 727 AACGCTGTTGATACAGCGCGCGGCGATGCTTTATCGCTGTTTTCGCGCATTTACTAC 786  
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263  
QY 787 GTCCAGCGCGGATGTGGAAGCGCCATGAAAGACCGTCTCTTTCGCGCTTTCAGC 846  
Db 264 AlaGluThrArgGluValValSerSerLeuLysLysAlaSerLeuTyrAlaAlaHisSer 283  
QY 847 GTCACCGGAAAGGCAACCAATCTCTTATCCAAGCATTTGAGCAA 891  
Db 284 IleThrArgProGlyThrGlnLysAlaTyrAlaSerIleAspGlu 298  
RESULT 7  
ADS25978  
ID ADS25978 standard; protein; 298 AA.  
XX  
AC ADS25978;  
XX  
DT 02-DEC-2004 (first entry)  
XX  
DE Bacterial polypeptide #15011.  
XX  
KW Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polypeptide.  
XX  
OS Bacteria.  
XX  
PN US2003233675-A1.  
XX  
PD 18-DEC-2003.  
XX  
PF 20-FEB-2003; 2003US-00369493.  
XX  
PR 21-FEB-2002; 2002US-0360039P.  
XX  
PA (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.

(CHEN/) CHEN X.  
(GOLD/) GOLDMAN B S.  
Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
WPI; 2004-061375/06.  
New recombinant DNA construct comprising a promoter positioned to provide  
for expression of a polynucleotide encoding a polypeptide from a  
microbial source, useful for producing plants with improved properties.  
Claim 1; SEQ ID NO 15011; 122pp; English.  
The invention relates to a recombinant DNA construct comprising a  
promoter functional in a plant cell, where the promoter is positioned to  
provide for expression of a polynucleotide encoding a polypeptide from a  
microbial source. The invention also relates to a transformed plant  
comprising the recombinant DNA construct and a method of producing a  
transformed plant having an improved property. The plant is a crop plant  
such as maize or soybean. The method of producing a transformed plant  
having an improved property comprises transforming a plant with the  
recombinant DNA construct and growing the transformed plant, where the  
polynucleotide or polypeptide is useful for improving plant properties.  
The recombinant DNA construct is useful for producing plants with  
improved plant properties, e.g. improved cold, heat or drought tolerance,  
tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
increased resistance to plant disease, better growth rate by modification  
of the cell cycle pathway with plant growth regulators, increased rate of  
homologous recombination, modified seed oil or protein yield and/or  
content, improved yield by modification of carbohydrate, nitrogen or  
phosphorus use and/or uptake, by modification of photosynthesis or by  
providing improved plant growth and development under at least one stress  
condition, improved lignin production or improved galactomannan  
production. This sequence represents a bacterial polypeptide used in the  
scope of the invention. Note: The sequence data for this patent did not  
form part of the printed specification but was obtained in electronic  
format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
XX  
SQ Sequence 298 AA;  
Alignment Scores:  
Pred. No.: 3 99e-89 Length: 298  
Score: 870.00 Matches: 167  
Percent Similarity: 76.61% Conservative: 59  
Best Local Similarity: 56.61% Mismatches: 69  
Query Match: 54.72% Indels: 0  
DB: 8 Gaps: 0  
US-10-049-750-11 (1-921) x ADS25978 (1-298)  
QY 7 ATCGCGGTTATTGGCTTAACATGGTGGACCTTATCACCAACCAAGATGCCCAA 66  
Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyValAspArgMetProGly 23  
QY 67 GAAGGGGAAACTCTGAAGCGCGCGGTTTAAATCGCTGCGCGGGAAGGGGGAAC 126  
Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43  
QY 127 CAGGCGCGCGCGCGCTAAGCTCAATTCAAAAGATTGATGTTGACCAAGTGGCGGAC 186  
Db 44 GlnAlaValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 63  
QY 187 GATATTTTTCGCGACACACACCATTCGTATCTCGAATCTCGGGGATCAATACGATAT 246  
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83  
QY 247 GTAGAAAAGTACCGTGATACAGCGCGCGTACCGCCGATTTTCTGTCACGCCCACTCC 306  
Db 84 ValValLysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103  
QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAAAGATTCTCTCGCGGGAAGATATCGAT 366  
Db 104 GluAenSerIleLeuIleValLysGlyAlaAenAlaAspLeuLeuProValGluValAsp 123



CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridization. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ffp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 321 AA;

Alignment Scores:  
 Pred. No.: 9.83e-85 Length: 321  
 Score: 831.50 Matches: 168  
 Percent Similarity: 73.97% Conservative: 48  
 Best Local Similarity: 57.53% Mismatches: 75  
 Query Match: 52.30% Indels: 1  
 DB: Gaps: 1

US-10-049-750-11 (1-921) x ABM57386 (1-321)

QY 1 ATGGATATCGCGTATTGGCTTAACATGCTGACCTTATACCTACCAACACGATG 60  
 DB 18 MetAspIleAlaValValGlySerAsnMetValAspLeuIleSerThrIleHisArgMet 37  
 QY 61 CCCAAAGAGGGGAAACTCTGGAAGCGCGCGCTTTAAATTCGGCTCGCGGAAAGGG 120  
 DB 38 ProSerAspGlyGluThrValGluAlaProGluPheArgMetGlyCysGlyGlyLeu 57  
 QY 121 GCGAACCGCGCTGCGCGCTAAGCTCAATCAAAAGTATGTATGTGACCAAGT 180  
 DB 58 AlaAsnGlnAlaValAlaAlaSerArgLeuGlyAlaGluValValMetValThrArgVal 77  
 QY 181 GCGGACGATATTTTCCCGCACACACATTCGTAATCTCGAATCCTCGGGGATCAATACG 240  
 DB 78 GlyAsnAspValPheAlaAspThrThrLeuAspAsnPheArgLysAsnGlyIleAspThr 97  
 QY 241 ACGTATGTAGAAAAGTACCGTGTACGAGCGCGGTAGCGCGGATTTTCGTCACAGCC 300  
 DB 98 ThrHisValLeuArgThrAsp---AlaSerSerGlyValAlaProIlePheValAspPro 116  
 QY 301 AACTCCAGCAACAGCATTCGTATCATCAAGGCGCTAACAAAGTTTCTCTCGCGGAGAT 360  
 DB 117 GluSerArgAsnSerIleLeuIleIleLysGlyAlaAsnAlaGlnLeuSerProAlaAsp 136  
 QY 361 ATCGATCGCGCGCGGAAAGATTTAAATAATGCGACGCTTATTTCTTCGCACTGGAGTT 420  
 DB 137 ValGluGlyAlaArgGlnAspIleAlaThrCysLysLeuIleMetLeuGlnLeuGluIle 156  
 QY 421 CAGCTTGAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACCGGATTTGAAGTGT 480  
 DB 157 ProLeuGluThrValThrAlaThrIleGluLeuGlyAspValLeuGlyIleProIleLeu 176  
 QY 481 TTAACCTCGCGCGCAGCATTTACGGAAATTAGATATATGCTTATCGCTGTAATCGGATTC 540  
 DB 177 LeuAsnProAlaProValAlaProGluLeuAspLeuGluArgIleArgGlyIleGluPhe 196  
 QY 541 TTGTACCTTAATGAACCGAGCTGGAAATATTAAACGGTATGCGGATGAGTATCTATGAC 600  
 DB 197 PheMetProAsnGluSerGluLeuGluLeuIleThrGlyMetProValAspThrLeuAsp 216  
 QY 601 CATATTCCGCGCAGCGCAGCTTCGCTGTTAGATAAAGGGCTGAACATAATATTATTCACC 660  
 DB 217 AspIleGlyLysAlaThrAspValLeuLeuGlyAlaGlyIleThrAsnIleIleValThr 236

QY 541 TTGTACCTTAATGAACCGAGCTGGAAATATTAAACGGTATGCGGATGAGTATCTATGAC 600  
 DB 197 PheMetProAsnGluSerGluLeuGluLeuIleThrGlyMetProValAspThrLeuAsp 216  
 QY 601 CATATTCCGCGCAGCGCAGCTTCGCTGTTAGATAAAGGGCTGAACATAATATTATTCACC 660  
 DB 217 AspIleGlyLysAlaThrAspValLeuLeuGlyAlaGlyIleThrAsnIleIleValThr 236  
 QY 661 ATGGCGGAGAAAGCGCGCTGTGGATGACGCGTGACCAAGATCCATGCTTCGCGGCTTT 720  
 DB 237 LeuGlySerArgGlyAlaMetTrpAlaHisAlaGluGlyArgLysIleIleLysAlaPro 256  
 QY 721 AGATGAACGCTTGTATACACGCGCGCGGCGATGCTTTATCGCGTGTTCGCGCAT 780  
 DB 257 ValValGlnAlaValAspThrThrGlyAlaGlyAspAlaPheIleGlyCysPheAlaLys 276  
 QY 781 TACTAGCTCAGACGCGGATGTGAAGCGCGCATGAAAGAAACCGCTCTCTTCGCGCT 840  
 DB 277 GluTrpValAspHisGlyAspValIleAlaAlaIleArgAlaGlyAsnArgTyrAlaAla 296  
 QY 841 TTCAGCTACACCGGAAAGGACCAATCCTCTTAT 876  
 DB 297 AspSerValThrArgHisGlyThrGlnSerSerTyr 308

RESULT 9  
 ABM57386  
 ID ABM57386 standard; protein; 321 AA.  
 XX  
 AC ABM57386;  
 XX  
 DT 20-OCT-2003 (first entry)  
 XX  
 DE Propionibacterium acnes predicted ORF-encoded polypeptide #22062.  
 XX  
 XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
 KW immunostimulant; immune response; vaccine.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WQ2003033515-A1.  
 XX  
 PD 24-APR-2003.  
 XX  
 PF 11-OCT-2002; 2002WO-US032727.  
 XX  
 PR 15-OCT-2001; 2001US-00978825.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Valliave-Douglas J;  
 XX  
 DR WPI; 2003-381789/36.  
 DR N-PSDB; ACF64541.  
 XX  
 PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX  
 XX Example 1; SEQ ID NO 22062; 1481pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,







QY 727 AACGCTGTTGATACACAGCGCGCGGCGATGCTTTATCGGCTGTTTCGGCGATTACTAC 786  
 Db 255 ThrProValAspThrThrGlyAlaGlyAspAlaPheIleGlyAlaPheAlaHisPheLeu 274  
 QY 787 GTCCAGAGCGGGGATGTGAAGCGCGCATGAAAAAGCCGCTCTCTTTGCGGCTTTTCAGC 846  
 Db 275 AlaAlaThrGlyGluValGluGlyAlaLeuAlaHisAlaAlaArgTyrAlaAlaHisSer 294  
 QY 847 GTCACCGGGAAGACGCCACCCCAATCTCTTAT 876  
 Db 295 ValThrGlyArgGlyThrGlnLysSerTyr 304.

## RESULT 11

AAU36261

ID AAU36261 standard; protein; 308 AA.

XX AC

XX AC

XX AC

DT 14-FEB-2002 (first entry)

XX

DE Pseudomonas aeruginosa cellular proliferation protein #251.

XX Antisense: prokaryotic cellular proliferation protein; antibiotic;

KW antibacterial; drug design.

KW

XX

OS Pseudomonas aeruginosa.

XX

XX WO200170955-A2.

PN

XX PD

XX 27-SEP-2001.

XX

PF 21-MAR-2001; 2001WO-US009180.

XX

XX 21-MAR-2000; 2000US-0191078P.

PR

PR 23-MAY-2000; 2000US-0206848P.

PR

PR 26-MAY-2000; 2000US-0207272P.

PR

PR 23-OCT-2000; 2000US-0242578P.

PR

PR 27-NOV-2000; 2000US-0253625P.

PR

PR 22-DEC-2000; 2000US-0257931P.

PR

PR 16-FEB-2001; 2001US-0269308P.

XX

XX (ELIT-) ELITRA PHARM INC.

PA

XX

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;

PI

PI Yamamoto RT, Xu HH;

XX

XX WPI; 2001-611495/70.

DR

DR N-PSDB; AAS54120.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

SQ Sequence 308 AA;

## Alignment Scores:

Pred. No.: 9.41e-53 Length: 308

Score: 550.00 Matches: 126

Percent Similarity: 59.33% Conservative: 52

Best Local Similarity: 42.00% Mismatches: 122

Query Match: 34.59% Indels: 0

DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x AAU36261 (1-308)

QY 7 ATCGCGGTTATTGGCTCTAACATGTTGGACCTTATCACCTACACCAACGAGATGCCAAA 66  
 Db 5 ValLeuValValGlySerLeuAenMetAspLeuValValArgAlaProArgLeuProArg 24  
 QY 67 GAAGGGAAACTCTGGAAGCGCGCTTTAAATCGGCTGGCGGCGGAAAGGGGGAAC 126  
 Db 25 GlyGlyGluThrLeuAlaGlyGlnSerPheIleThrValProGlyGlyGlyAlaAen 44  
 QY 127 CAGCGCGTGGCGCGCTTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCGGAC 186  
 Db 45 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValAlaMetIleGlyCysLeuGlyAsp 64  
 QY 187 GATATTTTGGCGCAACACCATTCGTAATCTCGAATCTGGGGATCAATACGACGTAT 246  
 Db 65 AspAlaTyrGlyAspGlnLeuTyrArgAlaLeuGlnAlaGluGlyIleAspCysGlnGly 84  
 QY 247 GTAGAAAAGTACCGGTGTACACGCGCGGTAGCGCCGATTTTCGTCAACGCCAACTCC 306  
 Db 85 ValGluArgValAlaGlyGlnSerSerGlyValAlaLeuIleValValAspAspSerSer 104  
 QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAAACAGTTCTCTCGCGGAAAGATATCGAT 366  
 Db 105 GlnAenAlaIleValIleValAlaGlyGlyAsnGlyHisLeuSerProAlaValLeuAla 124  
 QY 367 CGCGCGCGGAGAGATTTAAAAATCCAGCTTATTGTTCTGCAACTGGAGTTTCAGCTT 426  
 Db 125 ArgHisGluHisLeuLeuGluGlnAlaGlnValValValCysGlnLeuGluSerProLeu 144  
 QY 427 GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAACACGCGGATTCGAAGTGTATTAAAC 486  
 Db 145 GluThrValGlyHisValLeuArgArgAlaHisAlaLeuGlyLysThrValIleLeuAen 164  
 QY 487 CTGCGCCAGCATTTACGGAAATTAGATATGTTATGCTCTGCTGTAATATTCGATTCTTGTGA 546  
 Db 165 ProAlaProAlaThrArgAspValProAlaGluTrpLeuProLeuValAspTyrLeuVal 184  
 QY 547 CTTAATGAAACCGAGCTGAAATATTAACGGGTATGCCAGTGGATACCTATGACCATATT 606  
 Db 185 ProAenGluThrGluSerGluLeuLeuCysArgLeuProValAspSerLeuGluSerAla 204  
 QY 607 CGCGCAGCGGCGACGTTTCGCTGTAGATAAAGGCTGAACAATATTATTGTCACATGGGC 666  
 Db 205 GlyArgAlaAlaGluArgLeuArgGluMetGlyAlaGlyArgValIleValThrLeuGly 224  
 QY 667 GAGAAGGCGCGCTGTGGATGACCGCTGACAGGAGTCCATGTTCCCGCGTTTACAGTG 726  
 Db 225 AlaGlnGlyAlaLeuLeuValGlyGlyArgValGluHisPheProValAlaArgVal 244  
 QY 727 AACGCTGTTGATACAGCGCGCGCGGCGATGCTTATTCGGCTGTTTCGGCGATTACTAC 786  
 Db 245 LysAlaLeuAspThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaLeu 264  
 QY 787 GTCCAGCGGGGATGTGGAAGCGCGCATGAAAAAGCCGCTCTCTTCGCGCTTTTCAGC 846  
 Db 265 AlaArgGlyLeuAspGluAlaAlaAlaIleArgPheGlyGlnAlaAlaAlaIleSer 284  
 QY 847 GTCACCGGGAAGACGCCACCCCAATCTCTTATCCTCAAGCATTTGAGCAATTAATGATGATCT 906  
 Db 285 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluGluValGluArgAlaLeu 304

RESULT 12

New polynucleotides for the identification and development of antibiotics, comprise sequences of antisense nucleic acids.

Example 3; SEQ ID NO 11854; 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIFO at ftp.wifo.int/pub/published\_pct\_sequences







CC sequence.

XX Sequence 322 AA;

Search completed: December 23, 2005, 22:29:50  
Job time : 186.852 secs

Alignment Scores:

Pred. No.:	2,74e-52	Length:	322
Score:	546.00	Matches:	117
Percent Similarity:	58.98%	Conservative:	57
Best Local Similarity:	39.66%	Mismatches:	117
Query Match:	34.34%	Indels:	4
DB:	8	Gaps:	2

US-10-049-750-11 (1-921) x AQ089874 (1-322)

```
QY 7 ATCGCGGTATTGGCTTAACATGGTGGACCTTATCACCACCAACAGATGCCCAA 66
   :::::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 18 ValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProLys 37

QY 67 GAAGGGAACTCTGGAAGCGCGGCTTTAAATCGGCTGCGGCGGAAAGGGCGAAC 126
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 38 ThrGlyGluThrIleHisGlyHisLysPheIleGlyPheGlyGlyLysGlyAlaAsn 57

QY 127 CAGGCGCTGCGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGCAC 186
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 58 GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys 77

QY 187 GATATTTTTCGCGACAAACACCATTCGTAATCTCGAATCCTGGGGATCAATACGAGTAT 246
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 78 AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe 97

QY 247 GTAGAAAAGTACCGGTACACGACGCGGTAGCGCGGATTTTCGTCAACGCCCACTCC 306
   :::::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 98 ThrTyrGlnThrLysAspAlaAlaThrGlyThrAlaSerIleIleValAsnAsnGluGly 117

QY 307 AGCAACAGCATTCGATCATCAAGCGCTACAAAGTTCTCTCGCGGAAGATATCGAT 366
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 118 GlnAsnIleIleValIleValAlaGlyAlaAsnLeuLeuLeuAsnThrGluAspLeuArg 137

QY 367 CGCGCGCGGCGAAGATTAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAGTTCAGCTT 426
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 138 AlaAlaAlaAsnValIleSerArgAlaLysValMetValCysGlnLeuGluIleThrPro 157

QY 427 GAAACGGTTTATCACCAATAGAAATTTGGCAGAAACACAGGGATTCGAAGTGTATTAAAC 486
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 158 AlaThrSerLeuGluAlaLeuThrMetAlaArgSerGlyValLysThrLeuPheAsn 177

QY 487 CCTGCGCCAGCATTCAGGGAATTAGATATGCTTATGCGCTGTAATGCGATTTCTTTGTA 546
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 178 ProAlaProAlaIleAlaAspLeuAspProGlnPheTyrThrLeuSerAspValPheCys 197

QY 547 CCTAATGAAACCGAGTGGAAATATTAAACGGTATGCCAGTGGATACCTATGACCATATT 606
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 198 CysAsnGluSerGluAlaGluIleLeuThrGlyLeuThrValGlySerAlaAlaAspAla 217

QY 607 CCGCAGCGCGCAGTTCGCTGGTGTAGATAAAGGCTGAACAATATTATTCACCATGGGC 666
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 218 GlyGluAlaAlaLeuValLeuLysArgGlyCysGlnValValIleIleThrLeuGly 237

QY 667 GAGAAAGGC-----CGCTGTGGATGACCGCTGACCGAGGAGTCCATGTCGGCGTTT 720
   :::::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 238 AlaGluGlyCysValValLeuSerGlnThrGluProGluProLysHisIleProThrGlu 257

QY 721 AGAGTGAACGCTGTTGATACACGCGCGCGGCGATGCTTTTATCGGCTGTTTCGGCGCAT 780
   :::::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 258 LysValLysAlaValAspThrThrGlyAlaGlyAspSerPheValGlyAlaLeuAlaPhe 277

QY 781 TAC-----TACGTCCAGCGGGGATGTGGAAGCCCGCATGAAAGAAAGCCGCTCTTTT 834
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 278 TyrLeuAlaTyrTyrProAsnLeuSerLeuGluAspMetLeuAsnArgSerAsnPheIle 297

QY 835 GCGGCTTTCAGCGTCAACCGGGAAGGCACCCCAATCCTCTTATCCA 879
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 298 AlaAlaValSerValGlnAlaAlaGlyThrGlnSerSerTyrPro 312
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:17:41 ; Search time 37.391 Seconds  
(without alignments)  
4072.865 Million cell updates/sec

Title: US-10-049-750-11  
Perfect score: 1590  
Sequence: 1 atggatattcggtttattgg.....atctttctgtgaacgaataa 921

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlp  
-MODEL=framer\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO spool\_p/US10049750/runat\_23122005\_113553\_6317/app query fasta\_1.1742  
-DB=Issued\_Patents\_AA -QFWT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10049750 @CGN 1 1 101 @runat\_23122005\_113553\_6317 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5 COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6 COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1207	75.9	311	2	US-09-543-681A-5777
2	549	34.5	503	2	US-09-252-991A-31882
3	491.5	30.9	320	2	US-09-543-681A-5983
4	458	28.8	302	2	US-09-710-279-3062
5	452.5	28.5	330	2	US-09-489-039A-14043
6	452	28.4	315	2	US-09-134-001C-4521
7	447.5	28.1	311	2	US-09-543-681A-7498
8	420.5	26.4	406	2	US-09-543-681A-8124
9	387.5	24.4	342	2	US-09-489-039A-9668
10	377.5	23.7	328	2	US-09-489-039A-7290
11	376.5	23.7	529	2	US-09-134-000C-4524
12	310.5	19.5	451	2	US-09-543-681A-4841

13	250	15.7	170	2	US-09-107-532A-3946	Sequence 3946, Ap
14	234	14.7	319	2	US-09-710-279-786	Sequence 786, App
15	233	14.7	321	2	US-09-134-001C-3900	Sequence 3900, Ap
16	220.5	13.9	269	2	US-09-134-000C-5214	Sequence 5214, Ap
17	217.5	13.7	331	2	US-09-543-681A-7111	Sequence 7111, Ap
18	215	13.5	328	2	US-08-826-611-6	Sequence 6, Appli
19	215	13.5	362	2	US-09-328-352-7264	Sequence 7264, Ap
20	214.5	13.5	313	2	US-09-489-039A-12895	Sequence 12895, A
21	211.5	13.3	347	2	US-08-826-611-2	Sequence 2, Appli
22	198.5	12.5	324	2	US-09-489-039A-13558	Sequence 13558, A
23	194	12.2	308	2	US-08-913-816C-17	Sequence 17, Appl
24	191	12.0	372	2	US-08-918-249-2	Sequence 2, Appli
25	191	12.0	372	2	US-08-918-249-4	Sequence 4, Appli
26	191	12.0	372	2	US-09-345-603-2	Sequence 2, Appli
27	191	12.0	372	2	US-09-345-603-4	Sequence 4, Appli
28	189.5	11.9	330	2	US-08-936-165A-384	Sequence 384, App
29	189.5	11.9	331	2	US-09-902-540-10221	Sequence 10221, A
30	183.5	11.5	303	2	US-08-913-816C-2	Sequence 2, Appli
31	182.5	11.5	356	2	US-09-489-039A-12757	Sequence 12757, A
32	160	10.1	115	2	US-09-599-360B-81	Sequence 81, Appl
33	160	10.1	115	2	US-09-621-976-5617	Sequence 5617, Ap
34	160	10.1	341	2	US-09-489-039A-8625	Sequence 8625, Ap
35	156.5	9.8	314	2	US-09-489-039A-10377	Sequence 10377, A
36	154	9.7	324	2	US-09-328-352-7458	Sequence 7458, Ap
37	152.5	9.6	303	1	US-08-961-539-2	Sequence 2, Appli
38	152.5	9.6	303	2	US-09-185-826-2	Sequence 2, Appli
39	148.5	9.3	491	2	US-09-543-681A-6634	Sequence 6634, Ap
40	148	9.3	334	2	US-09-107-532A-6336	Sequence 6336, Ap
41	146.5	9.2	312	2	US-09-489-039A-8270	Sequence 8270, Ap
42	146.5	9.2	477	2	US-09-492-709A-325	Sequence 325, App
43	145.5	9.2	303	2	US-09-583-110-5072	Sequence 5072, Ap
44	145.5	9.2	311	2	US-09-107-433-4423	Sequence 4423, Ap
45	144	9.1	526	2	US-09-252-991A-32984	Sequence 32984, A

ALIGNMENTS

RESULT 1

US-09-543-681A-5777  
; Sequence 5777, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543, 681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5777  
; LENGTH: 311  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-5777

Alignment Scores:  
Pred. No.: 1.18e-130 Length: 311  
Score: 1207.00 Matches: 232  
Percent Similarity: 87.79% Conservative: 34  
Best Local Similarity: 76.57% Mismatches: 37  
Query Match: 75.91% Indels: 0  
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x US-09-543-681A-5777 (1-311)

QY 1 ATGGATATCGCGTTATTGGCTCTCAACATGTTGGACCTTATCACCTTACCAACACGATG 60  
Db 6 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyrlleAspArgMet 25  
QY 61 CCCAAGAGGGGAACCTCTGGAAGCGCGGCTTTAAATCGCTCGCGCGGAAAGGG 120

Db 26 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyLysGly 45  
QY 121 GGAAACCGCGCGTGGCGCCCTAAGCTCAATCAAAAGTATTGATGTTGACCAAGTG 180  
Db 46 AlaAenGlnAlaValAlaAlaAlaLysLeuAenSerLysValMetMetLeuThrLysVal 65  
QY 181 GCGCAGCATATTTTGGCCACACACCATCTCGAATCTCGAATCTTGGGGGATCAATACG 240  
Db 66 GlyAspAspIlePheAlaAspAnThrIleMetAenLeuGluSerTyrglyIleAnThr 85  
QY 241 ACCTATGTAGAAAAAGTACCGTGTACACAGCGCGGTAGCGCCGATTTTCGTCAACGCC 300  
Db 86 ArgTyrValGluLysValProGlyThrThrSerGlyValAlaProIlePheValThrSer 105  
QY 301 AACTCCAGCAACAGCATTTCTGATCATCAAGCGGCTTAACAAGTTTCTCTCGCGGAGAT 360  
Db 106 GlnSerSerAsnSerIleLeuIleValLysGlyAlaAenGlnHisLeuSerAlaGluAsp 125  
QY 361 ATCGATCGCGCGCGAAGATTAAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAGTT 420  
Db 126 IleAspArgAlaAlaValSerLeuLysLysCysArgMetIleValLeuGlnLeuGluIle 145  
QY 421 CAGCTTGAAACGGTTTATCACCAATAGAAATTTGGCAAGAAACACCGGGATTGAAGTTTA 480  
Db 146 ProLeuGluThrValTyrHisAlaIleGluPheGlyAsnLysHisIleProValLeu 165  
QY 481 TTAAACCTGCGCCAGCATTCAGGNAATTAGATATGTCTTATGCCGTGTAATTCGATTC 540  
Db 166 PheAenProAlaProAlaSerLysAlaLeuAspLeuThrIleAlaAlaGlnCysAspPhe 185  
QY 541 TTTGTACCTTAATGAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600  
Db 186 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProIleAnThrLeuAsp 205  
QY 601 CATATTCCGCGCAGCGCAGCTTCGTGGTAGATAAGGCTGAACAAATATTATGTCACC 660  
Db 206 GluIleArgGluAlaAlaTyrPheLeuLeuGluLysGlyPheLysAsnIleIleValThr 225  
QY 661 ATGGGGGAGAAAGCGCGCTGGATGACGCGTGACCGAGGAGTCCATGTTCCGGCGTTT 720  
Db 226 LeuGlyGluLysGlyAlaLeuTrpValAenGlyGluIleGluLysTyrlleProAlaIle 245  
QY 721 AGAGTGAACGCTGTTGATACACGCGCGCGGCGATGCTTATCGCTGTTTTCGGCGAT 780  
Db 246 GluValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 265  
QY 781 TACTACGTCAGAGCGGGATGTGGAAGCGCGCATGAAAAAGCCGCTCTCTTCCGCGT 840  
Db 266 TyrTyrValHisThrGlyAsnIleGluGluAlaLeuAsnLysAlaValMetPheSerGly 285  
QY 841 TTCAGCGTCACCGGGAAGCGACCCCAATCTCTTATCCAGCATTCAGCAATTTAATGAG 900  
Db 286 LeuSerValThrGlyLysGlyThrGlnSerSerTyProSerIleGluLupheSerGlu 305  
QY 901 TATCTTTTCG 909  
Db 306 PheLeuSer 308

## RESULT 2

US-09-252-991A-31882  
; Sequence 31882, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31882  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31882  
  
Alignment Scores:  
Pred. No.: 2,73e-54 Length: 503  
Score: 549.00 Matches: 126  
Percent Similarity: 59.14% Conservative: 52  
Best Local Similarity: 41.86% Mismatches: 123  
Query Match: 34.53% Indels: 0  
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x US-09-252-991A-31882 (1-503)

QY 7 ATFCGCGTTATTTGGCTTAACATGATGGTGAACCTTATCACTACCAACACAGATGCCAAA 66  
Db 200 ValLeuValValGlySerLeuAsnMetAspLeuValValArgAlaProArgLeuProArg 219  
QY 67 GAAGGGGAACTCTCGAAGCGCGCGTAAAAATCGGCTGCGCGGAAAGGGCGGAC 126  
Db 220 GlyGlyGluThrLeuAlaGlyGlnSerPheThrThrIleProGlyGlyLysGlyAlaAen 239  
QY 127 CAGGCGGTGGCGCGCTAAGCTCAATCAAAAGTATTGATGTTGACCAAAAGTGGCGAC 186  
Db 240 GlnAlaValAlaAlaAlaArgLeuGlyAlaGlyValAlaMetIleGlyCysLeuGlyAsp 259  
QY 187 GATATTTTGGCGACACACCATTCGTAAATCTCGAATCTCGGAGTCAATACGAGTAT 246  
Db 260 AspAlaTyrGlyAspGlnLeuTyrArgAlaLeuGlnAlaGluGlyIleAspCysGlnGly 279  
QY 247 GTAGAAAAGTACCGTGTACACAGCGCGGTAGCGCGATTTTCGTCAACGCCACTCC 306  
Db 280 ValGluArgValAlaGlyLysSerGlyValAlaLeuIleValValAspAspSerSer 299  
QY 307 AGCAACAGCATTCATCATCAAGCGCGCTAAACAGTTTCTCTCGCGGAGATATTCAT 366  
Db 300 GlnAenAlaIleValIleValAlaGlyGlyAsnGlyHisLeuSerProAlaValLeuAla 319  
QY 367 CCGCGCGCGGAAAGATTTAAAAAATGCCAGCTTATTTGTTCTGCAACTGGAAGTTCACT 426  
Db 320 ArgHisGluHisLeuLeuGluGlnAlaGlnValValValCysGlnLeuGluSerProLeu 339  
QY 427 GAAACGGTTTATCACCAATAGAAATTTGCAAGAAACACCGGATTCGAAGTGTATTAAAC 486  
Db 340 GluThrValGlyHisValLeuArgAlaHisAlaLeuGlyLysThrValIleLeuAen 359  
QY 487 CTGCGCCAGCATTTACGGAAATTAGATATGTCTTATGCTGTAAATGCGATTTCTTTGTA 546  
Db 360 ProAlaProAlaThrArgAspValProAlaAspTrpLeuProLeuValAspTyrLeuVal 379  
QY 547 CCTAATGAAACCGAGCTGGAATATTTAACCGGTATGCCAGTGCATACCTATGACCATATT 606  
Db 380 ProAsnGluThrGluSerGluLeuLysCysArgLeuProValAspSerLeuGluSerAla 399  
QY 607 CCGCGAGCGGCGAGTTCGCTGAGTAAAGGCTGAACAATATTATTTCTCACCATGGGC 666  
Db 400 GlyArgAlaAlaGluArgLeuArgGluMetGlyAlaGlyArgValIleValThrLeuGly 419  
QY 667 GAGAAAGCGCGCTGTGGATGACGCGTGACAGGAAGTCCATGTTCTCCGCGGTTTACGATG 726  
Db 420 AlaGlnGlyAlaLeuLeuValGlyGluGlyArgValGluHisPheProValAlaArgVal 439  
QY 727 AACCGTGTGTATCACAGCGCGCGCGATGCGCTTTATCGCTGTTTCCGCGATTTACTAC 786  
Db 440 LysAlaLeuAspThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaLeu 459  
QY 787 GTCCAGAGCGGGATGTGGAAGCGCCCATGAAAGCCGCTCTCTTTTCGCGCTTTTCAGC 846  
Db 460 AlaArgGlyLeuAspGluAlaAlaAlaIleArgPheGlyGlnAlaAlaAlaIleSer 479



```
QY 847 GTCACCGGAAAGGACCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGATGATCTT 906
Db 480 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluGluValGluArgAlaLeu 499
QY 907 TCG 909
Db 500 Ala 500
RESULT 3
US-09-543-681A-5983
; Sequence 5983, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5983
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5983
Alignment Scores:
Pred. No.: 9,5e-48 Length: 320
Score: 491.50 Matches: 114
Percent Similarity: 55.45% Conservative: 54
Best Local Similarity: 37.62% Mismatches: 130
Query Match: 30.91% Indels: 5
DB: 2 Gaps: 3
US-10-049-750-11 (1-921) x US-09-543-681A-5983 (1-320)
QY 7 ATCCGGTTATTGGCTTAACATGTTGGACCTTATCACCTACCAACAGATGCCCAA 66
Db 18 LeuAlaValLeuGlySerIleAsnValAspHisIleMetAsnIleAlaGlnPheProLys 37
QY 67 GAAGGGAAACTCTGAAGCGCGGCTTTAAATCGCTCGCGGGAAGGCGGAAC 126
Db 38 ProGlyGluThrValIleGlyHisAspTyrLysIleAlaPheGlyGlyLysGlyAlaAsn 57
QY 127 CAGGCGTGGCGCCGCTAAGCTCAATTCATAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 58 GlnAlaValAlaCysGlyArgSerGlyAlaAspIleThrPheIleAlaCysValGlyAsp 77
QY 187 GATATTTTTCGCGACAAACACCATTCGTAATCTCGAATCCTGGGGATCAATACGACGAT 246
Db 78 AspAlaIleGlyArgGluIleAlaGlnLeuLysThrAspAsnIleAspThrAspAla 97
QY 247 GTAGAAAAAGTACCGGTACACGACGCGGTAGCGCGGATTTTCGTCAACGCCCACTCC 306
Db 98 IleArgIleIleProLysThrProThrGlyValAlaMetIleLeuValAsnGluGlnGly 117
QY 307 ACACACGATCTCTGATCATCAAGCGCTAACAAGTTCTCTCGCGGGAAGATATCGAT 366
Db 118 GluAsnValIleSerIleValAlaGlyAlaAsnSerAlaLeuThrProSerHisLeuHis 137
QY 367 CCGCGCGGGAAGATTAAATAAATGCCAGCTTATTGTTCTGCAACTGGAAGTTCAGCTT 426
Db 138 GlnTyrArgHisIleIleGluGlnAlaAspAlaLeuLeuMetGlnLeuGluSerProLeu 157
QY 427 GAACCGTTTATCAGCAATAGAAATTTGGCAAGAACACGCGGATGAAAGTGTATTAAAC 486
Db 158 AspThrValPheGluAlaAlaLysGlnAlaLysAlaHisGlnThrLysValIleLeuAsn 177
QY 487 CTGCGCCAGCATTCAGGAAATAGATATGCTTATGCTGCTAAATGCGATTTCTTTGTA 546
Db 178 ProAlaProAla---GlnProLeuSerAspGluPheLeuSerPheIleAspIleThr 196
```

```
QY 547 CTAATAGAAACCGAGCTGGAAATATTTAACCGGTATGCCAGTGATACCTATGACCATATT 606
Db 197 ProAsnGluThrGluAlaGluIleLeuThrGlyIleSerVal-----HisAspGluVal 214
QY 607 CGCGCA-----CGCGCACGTTTCGGTGTAGATAAAGGCGTGAACAATATTATTGTGACC 660
Db 215 GlyAlaAlaLysAlaAlaAsnIleLeuHisSerLysGlyIleHisValLeuIleThr 234
QY 661 ATGGCGGAGAAAGCGCGCTGTGGATGACGCGTGACCGACGAGAAAGTCCATGTCGCGGTTT 720
Db 235 LeuGlySerArgIleValTrpPheSerGluGlnGlyThrGlyMetIleIleProGlyPhe 254
QY 721 AGAGTGAACGCTGTTGATACAGCGCGCGCGGATGCTTATTCGGCTGTTTCGCGCAT 780
Db 255 ArgValGluAlaValAspThrIleAlaAlaGlyAspThrPheAsnGlyAlaPheValThr 274
QY 781 TACTACGTCCAGACGCGGGATGTGGAAGCGCGCATGAAAGAAAGCCGCTCTTTGCGCGCT 840
Db 275 AlaIleLeuGluGlyLysSerAlaHisAspAlaIleArgPheAlaHisAlaAlaAla 294
QY 841 TTCAGCGTCACCGGGAAGGCAACCAATCCTCTTATCCAGCATTCAGCAATTTAATGAG 900
Db 295 IleAlaValThrArgHisGlyAlaGlnSerSerValProTrpArgAspGluIleLysSer 314
QY 901 TATCTTTTCG 909
Db 315 PheLeuAla 317
RESULT 4
US-09-710-279-3062
; Sequence 3062, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3062
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3062
Alignment Scores:
Pred. No.: 6,98e-44 Length: 302
Score: 458.00 Matches: 108
Percent Similarity: 55.96% Conservative: 61
Best Local Similarity: 35.76% Mismatches: 129
Query Match: 28.81% Indels: 4
DB: 2 Gaps: 4
US-10-049-750-11 (1-921) x US-09-710-279-3062 (1-302)
QY 7 ATCCGGTTATTGGCTTAACATGTTGGACCTTATCACCTACCAACAGATGCCCAA 66
Db 1 ValIleValIleGlySerThrAsnValAspLysPheLeuAsnValLysArgPheProLys 20
QY 67 GAAGGGAAACTCTGGAAGCG---CGCGGCTTTAAATCGCTGCGCGGAAAGGGCG 123
Db 21 ProGlyGluThrLeuHisIleAsnGlnAlaGlnLysGluPheGlyGlyLysGlyAla 40
QY 124 RACACGCGGTGGCGCGCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCG 183
Db 41 AsnGlnAlaIleAlaAlaSerArgLeuAlaAlaAspThrThrPheIleSerLysValGly 60
```

```
QY 184 GACGATATTTTCCGCAACACCACTTCGTAAATCTCGAATCTCTGGGGATCAATACGACG 243
Db 61 LysAspGlyAsnAlaAsnPhelLeuGluAspPheLysLeuAlaGlyAlaGlyHisThrGln 80
QY 244 TATGTAGAAAAGTAGTACCGGTGTACAGACGCGGTAGCGCCGATTTTGTCTCAACGCCAAC 303
Db 81 TyrIleLeuThrSerGluSerGluGluThrGlyGlnAlaPheIleThrValAspGluAla 100
QY 304 TCCAGCAACAGCATTTCTGATCATCAAAAGCGGTAAACAGTTTCTCTCGCGGGAAGATATC 363
Db 101 GlyGlnAsnThrIleLeuValTyrGlyGlyAlaAsnMetThrLeuSerAlaThrAspVal 120
QY 364 GATCGCGCGCGGAGAGATTAAAAATGCGCAGCTTATTTGTTCTGCACTGGAAGTTCAG 423
Db 121 GluMetSerAlaAspAlaPheIleGlyAlaAspPheValValAlaGlnLeuGluValPro 140
QY 424 CTTGAAACGGTTTATCACCAATAGAAATTTGGCAAGAAACACGCGGATTTGAAGTTATTA 483
Db 141 PheGluAlaIleGluGlnAlaPheLysIleAlaArgLysGlnAsnIleThrThrValLeu 160
QY 484 AACCTGCGCCAGCATTCGCGGAATTAGATATGCTTATGCTGCTGCTAAATGCGATTTCTTT 543
Db 161 AsnProAlaProAlaIle---GluLeuProLysSerLeuLeuGluLeuThrAspIleIle 179
QY 544 GTACCTAATGAACCGAGCTGGAATATTAACCGGTATGCCAGTGGATACCTATGACCAT 603
Db 180 IleProAsnGluThrGluAlaGluLeuLeuThrGlyIleSerIleAsnAsnGluSerAsp 199
QY 604 ATTGCGCAGCGCGCATTTTCGTGTGTAGATAAGGCGTCAACAATATTATTGTCCACCATG 663
Db 200 MetLysGluThrAlaThrTyrPheLeuAspLeuGlyIleSerAlaValleuIleThrLeu 219
QY 664 GCGAGAAAGCGCGCTGTGGATGACGCTGACAGGAAGTCCATGTTCCGCGGTTTAGA 723
Db 220 GlyGluGlnGlyThrTyrCysAlaTyrGlnGluGlnTyrLysMetIleProAlaCysAsn 239
QY 724 GTGACGCTGTGTATACACGCGCGCGGCGGTATTCGCTGTTGTTTC---GCGCAT 780
Db 240 ValLysAlaIleAspThrThrAlaAlaGlyAspThrPheIleGlyAlaPheLeuSerGlu 259
QY 781 TACTAGTCCAGAGCGGGATGTGGAAGCGCGCATGAAAAAGCGCTCTCTTTCGCGCT 840
Db 260 LeuAsnLysAspLeuSerAsnIleGluSerAlaIleArgLeuAlaAsnGlnAlaSerSer 279
QY 841 TTCAGCGTCAACCGGGAAGGCCCAATCTCTTATTCGAAGCATTCGACCAATTTAAT--- 897
Db 280 LeuThrValcIlnArgLysGlyAlaGlnAlaSerIleProThrArgLysgluValGluAla 299
QY 898 GAGTAT 903
Db 300 GluTyr 301
```

## RESULT 5

```
US-09-489-039A-14043
; Sequence 14043, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14043
; LENGTH: 330
; TYPE: PRN
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14043
```

```
Alignment Scores:
Pred. No.: 3,19e-43 Length: 330
Score: 452.50 Matches: 111
Percent Similarity: 53.31% Conservative: 50
Best Local Similarity: 36.75% Mismatches: 138
Query Match: 28.46% Indels: 3
DB: 2 Gaps: 3
```

US-10-049-750-11 (1-921) x US-09-489-039A-14043 (1-330)

```
QY 7 ATCGCGGTATTTGGCTCTAAACATGCTGGACCTTATCACCTACACCAACAGATGCCCAA 66
Db 28 LeuValValLeuGlySerIleAsnAlaAspHisIleLeuAsnLeuAspAlaPheProThr 47
QY 67 GAAGGGGAAACTCTGGAAGCGCGCGTGTAAATCGGCTGGCGGAAAGGGGCGAAC 126
Db 48 ProGlyGluThrValThrGlyHisIstyrGlnValAlaPheGlyGlyLysGlyAlaAsn 67
QY 127 CAGGCGGTGGCGCGCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 68 GlnAlaValAlaAlaGlyArgSerGlyAlaAspIleAlaPheIleAlaCysThrGlyAsp 87
QY 187 GATATTTTTCGCAACACCATTCGTAAATCTCGAATCTCTGGGATCTGAGGATCAATACGAT 246
Db 88 AspAspIleGlyGluArgIleArgGlnLeuAlaSerAspLysIleAspValAlaPro 107
QY 247 GTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGATTTTCGTCAACGCCCACTCC 306
Db 108 ValArgAlaValAlaGlyGluAlaThrGlyValAlaLeuIlePheValAsnAlaGluGly 127
QY 307 AGCAACAGCATTTCTCATCATCAAGCGCTAAACAAGTTTCTCTCGCGGAAGATATCGAT 366
Db 128 GluAsnValIleGlyIleHisAlaGlyAlaAsnAlaAlaLeuSerValSerGlnValGlu 147
QY 367 CGCGCGCGGAGAGATTTAAAAATGCGCAGCTTATTTGTTCTGCAACTGGAAGTTCAGTT 426
Db 148 AlaGluLysGluArgIleAlaSerAlaGlnAlaLeuLeuMetGlnLeuSerProLeu 167
QY 427 GAAACGGTTTATCACGCAATGAATTTGGCAAGAAACACGCGGATTTGAAGTGTATTAAAC 486
Db 168 GluSerValIleAlaAlaLysIleAlaHisHisHisThrThrValValLeuAsn 187
QY 487 CTGCGCGCAGCATTCGCGGAATTTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546
Db 188 ProAlaProAla---ArgGluLeuProAspGluLeuLeuAlaLeuValAspIleThr 206
QY 547 CTAATGAAACCGAGCTGGAATATTAACCGGTATGCCAGTGCAGTACCTATGACCATATT 606
Db 207 ProAsnGluThrGluAlaGluLysLeuThrGlyIleArgValGluSerAspGluAspAla 226
QY 607 CGCGCAGCGGACGTTTCGCTGTAGATAAAGCGCTGAACAATATTTATTCACCATGGGC 666
Db 227 AlaLysAlaAlaAspValLeuHisAlaLysGlyIleGlyThrValMetIleThrLeuGly 246
QY 667 GAGAAAGCGCGCTGTGGATGACGGGTGACCAAGGAGTC---CATGTTCCGGCGTTTGA 723
Db 247 SerArgGly---ValTrpLeuSerAlaGluGlyGluSerArgArgIleProGlyPheArg 265
QY 724 GTGAACGCTGTGTATACCAAGCGCGCGGCGATGCTTATTCGGCTGTGTTCCGCGCATAC 783
Db 266 ValGlnAlaIleAspThrIleAlaAlaGlyAspThrPheAsnGlyAlaLeuValThrAla 285
QY 784 TACTGCCAGCGGGGATGTGGAAGCGCGCATGAAAAAGCGCTCTCTTTCGCGCTTTC 843
Db 286 LeuLeuGluGlyThrAlaLeuProGluAlaIleArgPheAlaHisAlaAlaIle 305
QY 844 AGCGTCACCGGAAAGGCCAACCTCTTATTCGAAGCATTCGACCAATTAATAGTAT 903
Db 306 AlaValThrArgLysGlyAlaGlnProSerValProTrpArgThrGluIleAspGluPhe 325
QY 904 CTTTCG 909
Db 326 LeuAla 327
```

```

RESULT 6
US-09-134-001C-4521
; Sequence 4521, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4521
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4521

Alignment Scores:
Pred. No.: 3,55e-43 Length: 315
Score: 452.00 Matches: 107
Percent Similarity: 55.63% Conservative: 61
Best Local Similarity: 35.43% Mismatches: 130
Query Match: 28.43% Indels: 4
DB: 2 Gaps: 4

US-10-049-750-11 (1-921) x US-09-134-001C-4521 (1-315)
QY 7 ATCCGGTTATTGGCTTAACATGGTGGACCTTATCACCTACACCAACAGATGCCCAA 66
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
14 ValileValileGlySerThrAsnValAspLysPheLeuAsnVallyAspGlyPheProLys 33
QY 67 GAAGGGGAAACTCTGGAAGCG---CCGGCGTTTAAATCGGCTGCGCGGAAAGGGCG 123
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 ProGlyGluThrLeuHisIleAsnGlnAlaGlnLysGluPheGlyGlyGlyGlyAla 53
QY 124 AACGAGCGCTGGCGCCCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCG 183
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 AsnGlnAlaIleAlaAlaSerArgLeuAlaAlaAspThrPheIleSerLysValGly 73
QY 184 GACGATATTTTCCGACAAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACG 243
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
74 LysAspGlyAsnAlaAsnPhelleLeuGluAspPheLysLysAlaGlyIleHisThrGln 93
QY 244 TATGTAGAAAAAGTACCGTGTACACGACGGCGGTAGCGCGGATTTTCTCAACGCCAAC 303
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
94 TyrIleLeuThrSerGluSerGluGluThrGlyGlnAlaPheIleThrValAspGluAla 113
QY 304 TCAGCAACAGATTTCTGATCATCAAGCGCTAACAAGTTTCTCTCGCGGGAAGATATC 363
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
114 GlyGlnAsnThrIleLeuValTyrGlyGlyAlaAsnMetThrLeuSerAlaThrAspVal 133
QY 364 GATCGCGCGCGGAACATTTAAAAAATCCAGCTTATTGTTCTGCAACTGGAAGTTGAG 423
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 GluMetSerValAspAlaPheIleGlyAlaAspPheValValAlaGlnLeuGluValPro 153
QY 424 CTTGAACCGTTTATCACGCAATAGAAATTTGGCAAGAAACACAGGGATGAAGTTGTTATTA 483
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 PheGluAlaIleGluGlnAlaPheLysIleAlaAatGlyGlnAsnIleThrThrValLeu 173
QY 484 AACCTGCGCGCAGCATTCAGGAATATAGATATGCTTATGCTGCTGTAATTCGATTTCTTT 543
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
174 AsnProAlaProAlaIle---GluLeuProLysSerSerLeuGluLeuThrAspIleIle 192
QY 544 GTACCTAATGAACCGAGCTGGAATATTAAACCGGTATCCAGTATGATACCTATGACCAT 603
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
193 IleProAsnGluThrGluAlaGluLeuLeuThrGlyIleSerIleAsnAsnGluSerAsp 212
```

```

QY 604 ATTCGGCGACGGCAGCTTCGCTGTAGATAAAGGCTGAACAATATTATTGTCACCATG 663
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
213 MetLysGluThrAlaThrThrPheLeuAspLeuGlyIleSerAlaValLeuIleThrLeu 232
QY 664 GCGGAGAAAGCGCGCTGTGGATGACGCGTACGAGGAAGTCCATGTCGGGGTTTGA 723
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
233 GlyGluGlnGlyThrTyrCysAlaTyrGlnGluGlnTyrLysMetIleProAlaCysAsn 252
QY 724 GTGAACGCTGTTGATACACGCGCGCGCGATGCTTATCGGCTGTTTC---GCGCAT 780
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
253 ValLysAlaIleAspThrThrAlaAlaGlyAspThrPheIleGlyAlaPheLeuSerGlu 272
QY 781 TACTAGCTCACAGCGGGGATGTGNAAGCGGCATGAAAGCCCTCTCTTGGCGCT 840
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 LeuAsnLysAspLeuSerAsnLeuGluSerAlaIleArgLeuAlaAsnGlnAlaSerSer 292
QY 841 TTCAGCGTCACCGGAAAGGACCAATCTCTTATCCAGCATTCAGCAATTAAT--- 897
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
293 LeuThrValGlnArgLysGlyAlaGlnAlaSerIleProThrArgLysGluValGluAla 312
QY 898 GAGTAT 903
Db : |||||
313 GluTyr 314

RESULT 7
US-09-543-681A-7498
; Sequence 7498, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7498
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7498

Alignment Scores:
Pred. No.: 1.17e-42 Length: 311
Score: 447.50 Matches: 108
Percent Similarity: 55.08% Conservative: 60
Best Local Similarity: 35.41% Mismatches: 134
Query Match: 28.14% Indels: 3
DB: 2 Gaps: 3

US-10-049-750-11 (1-921) x US-09-543-681A-7498 (1-311)
QY 7 ATCCGGTTATTGGCTTAACATGGTGGACCTTATCACCTACACCAACAGATGCCCAA 66
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
9 LeuValIleLeuGlySerValAsnValAspHisIleLeuAsnValAlaHisPheProHis 28
QY 67 GAAGGGGAAACTCTGGAAGCGCGGCTTTAAATTCGGCTGCGGGGAAAGGGGCGAAC 126
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29 ProGlyGluThrIleSerGlyAsnGlnTyrGlnMetValPheGlyGlyGlyAlaAsn 48
QY 127 CAGGCGCTGGCGCGCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 GlnAlaValAlaAlaGlyArgCysGlyAlaAsnIleThrPheLeuAlaCysLeuGlyAsn 68
QY 187 GATATTTTTCGCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGAGTAT 246
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
69 AspAspIleGlyGlnSerAlaLysAlaGlnLeuIleThrAspLysIleAspThrAspCys 88
QY 247 GTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGATTTTCGTCAACGCCAATCC 306
Db : ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
89 IleGluLeuIleAspAspGluAlaThrGlyValAlaLeuIlePheValAsnGlnGlnGly 108
```

QY 307 AGCAACAGCATTCATCATCAAAAGCGCGTAAACAAGTTTCTCTCGCGGAGATATCGAT 366  
DB 109 GluAenValIleGlyIleHisAlaGlyAlaAenGlyArgLeuAenArgGlyTyrValGlu 128  
QY 367 CGCGCGGCGGAGATTAAATAAATGCCAGCTATTGTTCTGCAACTGGAAGTTCCAGCTT 426  
DB 129 ArgTyrAenSerValIleLysLysAlaAenAlaLeuLeuMetGlnLeuGluSerProLeu 148  
QY 427 GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAACACCGGATGGAAGTGTATTAAC 486  
DB 149 AspSerValLeuLysAlaAlaGlnValAlaLysGlnGluSerValGlnValIleLeuAen 168  
QY 487 CCGCGCCAGCATTAAGGAATTAGATATGCTTATGCTGTAATGCGATTCTTTGTA 546  
DB 169 ProAlaProAla---GlnAlaLeuProAenGluLeuLeuSerLeuValAspIleThr 187  
QY 547 CTAATGAAACCGAGCTGGAATATTAACCGGTATGCCAGTGGATACCTATGACCATATT 606  
DB 188 ProAenGluThrGluThrGluTyrLeuThrGlyValLysValIleAenAspGluSerAla 207  
QY 607 CGCGCAGCGGCGAGTTCGCTGTAGATAAAGCGGTGAACAATATTATGTCACCATGGGC 666  
DB 208 GlnArgAlaSerAsnIleLeuHisAspLysGlyIleGluThrValLeuIleThrLeuGly 227  
QY 667 GAGAAAGCGCGCTGGTATGACGCGT---GACCAGGAAGTCCATGTTCCGGCGTTTGA 723  
DB 228 SerArgGly---ValTrpValSerGlnAenAenGlnGlyThrMetValProAlaPheLys 246  
QY 724 GTGAACGCTGTTGATACCAAGCGCGCGCGGATGCTTTATCGGCTGTTTCGCGCATTAC 783  
DB 247 ValThrAlaIleAspThrIleAlaAlaGlyAspThrPheAenGlyAlaPheIleThrAla 266  
QY 784 TAGCTCCAGAGCGGGATGTGAAGCGCGCATGAAAGCGCTCTCTTTCGCGCTTTC 843  
DB 267 LeuLeuGluGlyLeuProMetLysAlaIleLysPheAlaHisAlaAlaAlaIle 286  
QY 844 ACCTGACCGGGAAGGACCCCAATCTCTTATCCAAAGCATTTGACCAATTTAATGAGTAT 903  
DB 287 AlaValThrArgAlaGlyAlaGlnProSerValProTrpArgHisGluIleGluAlaPhe 306  
QY 904 CTTTCGTTGAACGAA 918  
DB 307 LeuSerGlnHisAsp 311

## RESULT 8

US-09-543-681A-8124  
; Sequence 8124, Application US/09543681A  
; Patent No. 6605709

## GENERAL INFORMATION:

; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 8124

; LENGTH: 406

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-8124

## Alignment Scores:

Pred. No.:	1,85e-39	Length:	406
Score:	420.50	Matches:	98
Percent Similarity:	54.43%	Conservative:	68
Best Local Similarity:	32.13%	Mismatches:	134
Query Match:	26.45%	Indels:	5
DB:	2	Gaps:	3

US-10-049-750-11 (1-921) x US-09-543-681A-8124 (1-406)

QY 7 ATCCGGGTATTGGCTTAACATGCTGACCTTATCACTACCAACACAGATGCCCAA 66  
DB 94 ValCysValLeuGlySerPheAenValAspMetIleSerTyrLeuProArgLeuProGlu 113  
QY 67 GAAGGGGAAACTCTGGAAGCGCGCGTAAATCGGCTCGCGGGAAGGGCGAAG 126  
DB 114 ThrGlyGluSerLeuAlaAenLysPheIlePheSerProGlyGlyGlyCysAen 133  
QY 127 CAGCGCTGGCGCGCTAAGCTCAATTCAAAGTATTGATGTTTACCAAGTGGCGAC 186  
DB 134 GlnAlaLeuAlaSerTyrAlaAenValHisPheIleThrLysIleGlyThr 153  
QY 187 GATATTTTTCGCGACACACCATTCGTAATCTCGAATCTCGGGATCAATACGACGTAT 246  
DB 154 AspGlnPheSerAspTyrAlaValAenPheIleSerSerArgIleLysThrSerThr 173  
QY 247 GTAGAAAAAGTACCGTGTACACGACGCGCGTAGCGCGGATTTTCGTCAACGCCAATCC 306  
DB 174 IleTyrGlnThrGluAenHisGlnThrGlyThrAlaSerIlePheValSerGluGluSer 193  
QY 307 AGCAACAGCATTCG---ATCATCAAGCGCTAACAAGTTTCTCTCGCGGGAAGATATC 363  
DB 194 GlyGluAenIleSerIleTyrSerGlyAlaAenMetAspIleSerSerAspGluVal 213  
QY 364 GATCCGCGCGGAGAGATTTAAAAAATGCCAGCTATTGTTCTGCAACTGGAAGTTGAG 423  
DB 214 LysIleGlnLysAspLysIleIleAspAlaAspIleIleLeuLeuGlnLeuIleThrAen 233  
QY 424 CTTGAAACGGTTTATCAGCCCAATAGAAATTTGGCAAGAACACGCGGATGGAAGTATTAT 483  
DB 234 IleGluAlaLeuLysGluIleSerIleGlyAenGluAenAenIleProIleIleLeu 253  
QY 484 AACCTCGCGCCAGCATTAAGGGAATTAGATGCTTATGCTGCTGTAATTCGATTTCTTT 543  
DB 254 AsnProAlaProTyrAsnLysIleIleAsp---LysLeuLeuProMetLeuAspIleLeu 272  
QY 544 GTACCTAATGAACCGAGCTGGAATATTAACCGGTATGCCAGTGGATACCTATGACCAT 603  
DB 273 ThrProAenGluThrGluAlaSerLeuLeuSerGlyIleGluValValAspLeuValSer 292  
QY 604 ATTCGCGCAGCGGACCGTTCGCTGTAGATAAAGCGCTGAAACAATATTATTCACCATG 663  
DB 293 AlaLysAenAlaAlaAenAlaIleTyrGlnLysGlyValAenLysValIleIleThrLeu 312  
QY 664 GCGGAGAAAGCGCGCTGTGGATGAGCGGTGACAGAGAGTCCATGTTCCGCGGTTTGA 723  
DB 313 GlySerLysGlySerLeuAlaTyrAspGlyTyrLysTyrIleTyrSerProAlaTyrPro 332  
QY 724 GTGAACGCTGTTGATACCAAGCGCGCGGATGCTTTATCGGCTGTTTCGCGCATTAC 783  
DB 333 AlaValValLysAenThrAlaGlyAlaGlyAspAlaPheAenGlyAlaLeuAlaSer 352  
QY 784 TACGTCCAGAGCGGGATGTGAAGCGCGCATGAAAGCGCTCTCTTTCGCGCTTTC 843  
DB 353 LeuAlaLysGlyLysGlnPheSerTyrAlaLeuArgTyrAlaSerAlaPheAlaSerLeu 372  
QY 844 AGCGTCACCGGGAAGGACCCCAATCTCTTATCCAAAGCATTTGACCAATTTAATGAGTAT 903  
DB 373 AlaValGlu-----ThrSerAenAlaSerGluMetProGluAspIleAenValMet 389  
QY 904 CTTTCGTTGAACGAA 918  
DB 390 HisArgIleAenGln 394

## RESULT 9

US-09-489-039A-9668  
; Sequence 9668, Application US/09489039A  
; Patent No. 6610836

## GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 2709.2004001  
;; CURRENT APPLICATION NUMBER: US/09/489,039A  
;; CURRENT FILING DATE: 2000-01-27  
;; PRIOR APPLICATION NUMBER: US 60/117,747  
;; PRIOR FILING DATE: 1999-01-29  
;; NUMBER OF SEQ ID NOS: 14342  
;; SEQ ID NO 9668  
;; LENGTH: 342  
;; TYPE: PRT  
;; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9668

Alignment Scores:  
Pred. No.: 1,111-35 Length: 342  
Score: 387.50 Matches: 103  
Percent Similarity: 50.16% Conservative: 50  
Best Local Similarity: 33.77% Mismatches: 143  
Query Match: 24.37% Indels: 9  
DB: 2 Gaps: 4

US-10-049-750-11 (1-921) x US-09-489-039A-9668 (1-342)

QY 7 ATCCGGTTATTGGCTCTAACATGGTGGACCTTATCACCTACACCAACAGATGCCCAA 66  
Db 34 ValCysValPheGlySerPheAsnPheAspMetValAlaArgValAspArgPheProVal 53  
QY 67 GAAGGGGAAACTCTGAAGCGCGCGGCTTTAAATCGGCTCGCGGAAAGGGGCGAAC 126  
Db 54 ProGlyGluSerLeuValAlaCysGlySerMetThrSerAlaGlyGlyGlyAlaAsn 73  
QY 127 CAGCGCGTGGCGCCCTGAAGCTCAATTCAAAGTATTGATGTGACCAAGTGGCGAC 186  
Db 74 GlnAlaThrAlaAlaLeuLysAlaGlyAlaAsnValHisTyrIleGlyLysIleGlyAsn 93  
QY 187 GATATTTTCCGACACACCATTCGTAATCTCGAATCTGGGATCAATACGAGTAT 246  
Db 94 AspThrPheGlyHisPheAlaArgArgHisLeuLysGlyValGlyPheAsnAlaValThr 113  
QY 247 GTAGAAAAAGTACCGGTGTACCGACGCGGTAGCGCGATTTTCGTCAACGCCAAC--- 303  
Db 114 LeuLeuValAlaGluThrProThrGlyAsnAlaLeuIleTyrValAlaGlyAsnAsp 133  
QY 304 TCCAGCAACGACATTCGATCATCAAGGCGCTAACAAAGTTCTCTCGCGGGAAGATATC 363  
Db 134 AlaGluAsnMetIleAlaValAspProGlyAlaAsnMetThrValThrAspAspGluIle 153  
QY 364 GATCGCGCGGGAAGATTTAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAGTTCAG 423  
Db 154 AlaGlyCysIleProAlaIleGlyCysAlaAspValValLeuValGlnLeuGluAsnAsn 173  
QY 424 CTTGAAACCGTTTATCACGCAATAGAAATTTGGCAAGAAACACGGGATTGAAGTGTATTAT 483  
Db 174 LeuSerAlaIleGluGlnValIleAspAlaGlyLysGlnAlaGlyAlaLeuValIleLeu 193  
QY 484 AACCTCGCCAGCATTCACGGGAATTAGATATGCTTATGCTGCTGTAATTCGCGATTCTTT 543  
Db 194 AsnProAlaPro---TrpGlnProValGluHisAlaLeuLeuArgLysValAspLeuLeu 212  
QY 544 GTACCTAATCAACCGAGCTGGAATATTAAACGGTATGCCAGTGCATACCTATGACCAT 603  
Db 213 ThrProAsnAlaThrGluAlaGlyLeuMetThrGlyArgValAspSerLeuThrAla 232  
QY 604 ATTCGCGCAGCGCAGCTTCGCTGGTAGATAAGGCTGAAACATATTATTGTCACCATG 663  
Db 233 AlaAlaGluAlaAlaAspValLeuHisAlaGlnGlyAlaArgAsnValIleThrLeu 252  
QY 664 GCGGAGAAAGCGCGCTGGATGACGCGTGACAGGAAGTCCATGCTCCGCGGTTTGA 723  
Db 253 GlyAlaSerGlyAlaLeuLeuSerGluHisGlyValLysSerProLysPhePro 272  
QY 724 GTGAACGCTGTGATACAGCGCGCGCGGATGCTTATTCGCTGTTTCGGCGATTAC 783  
::: |||||:::|||||

Db 273 SerHisProArgAspThrThrGlyAlaGlyAspAlaPheAsnGlyAlaLeuAlaAArg 292  
QY 784 TAGCTCCAGAGCGGGATGTGAAGCGCATGAAAGAAAGCGCTCTCTTTCCGCTTTC 843  
Db 293 LeuAlaCysGlyGluProLeuGlnAlaAlaAArgPheAlaAlaAArgVal 312  
QY 844 AGCGTCACCGGGAAGGACCAACCTCTTATCCAGCATTCAGCAATTAATGAGTAT 903  
Db 313 SerValGluLysGlnGlyAla---SerSerLeuPro-----GluTyr 325  
QY 904 CTTTCGTTGAACGAA 918  
Db 326 LeuGluAlaGlnGlu 330

#### RESULT 10

US-09-489-039A-7290  
; Sequence 7290, Application US/09489039A  
; Patent No. 6610836

#### GENERAL INFORMATION:

;; APPLICANT: Gary Breton et. al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
;; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 2709.2004001  
;; CURRENT APPLICATION NUMBER: US/09/489,039A  
;; CURRENT FILING DATE: 2000-01-27  
;; PRIOR APPLICATION NUMBER: US 60/117,747  
;; PRIOR FILING DATE: 1999-01-29  
;; NUMBER OF SEQ ID NOS: 14342  
;; SEQ ID NO 7290  
;; LENGTH: 328  
;; TYPE: PRT  
;; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-7290

#### Alignment Scores:

Pred. No.: 1,561-34 Length: 328  
Score: 377.50 Matches: 99  
Percent Similarity: 51.16% Conservative: 56  
Best Local Similarity: 32.67% Mismatches: 141  
Query Match: 23.74% Indels: 7  
DB: 2 Gaps: 5

US-10-049-750-11 (1-921) x US-09-489-039A-7290 (1-328)

QY 4 GATATCGCGTTATTGGCTCTAACATGGTGGACCTTATCACCTACACCAACAGATGCC 63  
Db 24 AspValValIleGlySerLeuAsnTyrAspIleLeuValGlnGlnAspArgLeuPro 43  
QY 64 AAAGAAGGGGAACTCTGGAAGCGCGGCTTTAAATCGGCTCGCGGAAAGGGCG 123  
Db 44 GluLeuGlyGluThrPheThrGlyAsnGluLeuMetLeuMetProGlyGlyValAla 63  
QY 124 AACCAAGCGCTCGCGCGCTTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGC 183  
Db 64 AsnGlnAlaValGlnCysSerArgLeuGlyLeuAsnValSerMetValGlyCysValGly 83  
QY 184 GACGATATTTTCCGACCAACACCATTCGTAATCTCGAATCTCGGGGATCAATCAGC 243  
Db 84 AsnAspIleTyrGlySerGluLeuIleAsnSerLeuArgGluAsnValSerValGlu 103  
QY 244 TATGTAGAAAAGTACCGGTGTACCGACGCGCGGTAGCGCGATTTTCGTCAACGCCAC 303  
Db 104 AsnValAsnLys---ArgGlyThrThrGlyIleGlyIleValGlnIleLeuGluSer 122  
QY 304 TCCAGCAACGACATTCGATCATCAAGGCGCTTAACAAAGTTCCTCTCGCGGGAAGATATC 363  
Db 123 GlyAspTyrCysSerThrIleIleLysGlyAlaAsnTyrLeuIleSerGluAspAspIle 142  
QY 364 GATCGCGCGGGAAGATTTAAAAAATGCCAG---CTTATTGTTCTGCAACTGGAAGTT 420  
Db 143 -----ThrGluSerLeuPheGluGlyGlnProLeuValIleLeuGlnSerGluIle 159  
QY 421 CAGCTTGAAACCGTTTATCACCAATAGAAATTTGGCAAGAAACACGGGATGAGTGTTA 480

	160	ProAlaProValGluTyrIleileGlyValAlaSerThrHisHisCysArgIleIle	179
D <sub>b</sub>		:::         ::::	
Q <sub>y</sub>	481	T <sub>TTAAACCCCTGC</sub> GCGCAGCATTTACGGGAATTAGATATGTCTTATGCCGTGTAATAATGCCATTTC	540
D <sub>b</sub>	180	LeuAsnAenAlaProAla---ArgaspValSerAlaHisAlaLeuSerLeuValAspTyr	198
Q <sub>y</sub>	541	T <sub>TTTGTTACCTAATGAAC</sub> CGAGCTGGAAAATATTAAACCGGTATGTCAGTGCATACCTATGCAC	600
D <sub>b</sub>	199	LeuValValAsnGluThrGluAlaAlaPheMetSerGlyAlaAspValSerSerIleAsp	218
Q <sub>y</sub>	601	CATATTCGCGGAGCGGCACGTTCCGTCGTAGATAAAGCGCTCAACAATATTATTGTCACC	660
D <sub>b</sub>	219	AspAlaHisSerCysAlaThrGlyLeuHisLysArgValLysGlyGlnValIleIleThr	238
Q <sub>y</sub>	661	ATGGCGCAGAAAGCGCGCTCTGGATGATCGCGTGACACGGAAGTCCATGTTCCCGCGC---	717
D <sub>b</sub>	239	LeuGlyGluLysGlyAlaValLeuSerGlnGluHisGlyThrArgHisPheProAlaVal	258
Q <sub>y</sub>	718	T <sub>TTTAGATGAACGCTGTT</sub> GTATACACAGCGCGCGCGCGATGCCCTTTATCGCTGTGTTTCGGG	777
D <sub>b</sub>	259	PheCysProAspValValAspThrGlyAlaGlyAspSerPheIleGlyIleAla	278
Q <sub>y</sub>	778	CATTACTACGTCACGAGCGGGGATGTGGAAGCGGCATGAAAAAGCGTCCTCTTTGGC	837
D <sub>b</sub>	279	TyrCysIleValAsnGlyPheSerLeuThrAspAlaIlePropheAlaAlaGluIleSer	298
Q <sub>y</sub>	838	GCTTTTCAGCGTCACCGGAAAGGCCCAATCTCTTATCCAAGCATGTAGCAATTTAAT	897
D <sub>b</sub>	299	SerCysSerIleGlnLysTyrglyGlyGlnAenSerPheProMetLeuProAspValSer	318
Q <sub>y</sub>	898	GAGTATCTT	906
D <sub>b</sub>	319	HisAlaLeu	321

```

RESULT 11
US-09-134-000C-4524
; Sequence 4524, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4524
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-134-000C-4524

```

QY 133 GTGCGCCGGCTAAGCTCAATTCAAAGATATTGTGATTTTGACCAAAAGTGGCGCAGCATATT 192  
||| ||||| :  
Db 261 ValglyLeuThrLysLeuGlyHisGlnAlaThrLeuIleGlyCysLeuGlySerAspThr 280  
||| ||||| :  
QY 193 TTTGCCGACAACACCATTTCGAATCTCGAATCCTGGGGGATCAATACGACGTATGTAGA 252  
||||| :  
Db 281 AspAlaAsnTyrlLeutyrsGluLeuGluLysTyrlHisValThrThrAspGlylleThr 300  
||||| :  
QY 253 AAAGTACCGTGTACCAGCAGCGGGGTAGCGCGATTTTCGTCAACGCCAACATCCAGCAAC 312  
:  
Db 301 ArgileGlnAspThrGluThrGlyGlnAlatyrlleTyrlValGluThrSerGlyaspSer 320  
||| ||||| :  
QY 313 AGCATTTCTGTATCATCAAGCGCGCTAAACAAGTTTCTCTCGCGGAAGATATCGATCGCGG 372  
||||| :  
Db 321 MetileSerilleLeuProGlyAlaAsnThrAlaLeuThrProLysLysilleAlaGlnGln 340  
||||| :  
QY 373 GCGGAAGATTTAAAAAATGCCAGCTTATTTGTTCTGCGAACTGGGAAGTTTCAGCTTGAACG 432  
:  
Db 341 LysHisLeuPheMetAspAlaSerPheCysLeuilleGlnThrGluIleProleuSerAla 360  
||| ||||| :  
QY 433 GTTTTATCACGCATATAGAAATTTGGCAAGAAAACACAGGGATTGAAGTGTATTATAACCTGGG 492  
||| ||||| :  
Db 361 ValGluLysAlaCysGluIleAlaGlnHisSerGlyValProIleileLeuLysProAla 380  
||||| :  
QY 493 CCAGCATTTACGGGAATTAGATATGTTCTTATGCTTATTAATGCGATTCTTTCTGTACTAAT 552  
||| :  
Db 381 --AlaIleHisIleProValAsnIleLeuGluLysValaspPhePheIleProAsn 399  
||| :  
QY 553 GAACCCGAGCTGGAAATATTA-----ACCGTATGCCAGTGGATACCTATGACCAT 603  
||| ||||| :  
Db 400 GluAspGluLeuLeuGluLeuGlnProAspThrGlyThr----- 412  
||| ||||| :  
QY 604 ATTCGCGCAGCGGACGCTTCGCTGTAGATAAAGGCTGAACAATATTATTGTCACCATG 663  
:  
Db 413 LeuGluGluLysAlaAlaTyrlPheLeuGluMetGlyValLysAsnValleValThrLeu 432  
||| :  
QY 664 GCGGAGAAAGCGCGCTGTGATGACGCGTGACACGGAAGTCTCTTTCGCGGTTTAGA 723  
||| ||||| :  
Db 433 GlyLysLysGlyValLeuLeuLysThrProGlnValCysHisTyrlPheProAlaThrGlu 452  
||| ||||| :  
QY 724 GTGAACGCTGTTGTATACACGCGCGCGCGGATGCTTTATCGCGCTGTTTCGCGCATTAC 783  
||| :  
Db 453 AsnIleAlaValaspSerThrGlyAlaSeraspSerPheIleSerAlaLeuAlaSerTyrl 472  
||| :  
QY 784 TAGTCCAGACGGGGATGTGGAAGCCCGCATGAAAAAGCGCTCTCTTTCCGCTTTC 843  
:  
Db 473 LeuSerLysGlyTyrlProThrGluAlaAlaIleGlnIleAlaIleAlaGlyPhe 492  
||| :  
QY 844 AGCGTCCACGGGAAAGGCACCAATCTCTTATCCAAGCATTTAGCAATTTAATGAGTAT 903  
||| ||||| :  
Db 493 SerValSerLysGluGlyValleaspSerLeuValaspHisValThrLeuGluAsnTyrl 512  
||| :  
QY 904 CTTTCGTTGAACGAA 918  
||| ||||| :  
Db 513 LeuIleLysLysGlu 517  
||| ||||| :  
  
RESULT 12  
US-09-543-681A-4841  
; Sequence 4841, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTE  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4841  
; LENGTH: 451

```
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-4841

Alignment Scores:
Pred. No.: 11e-26 Length: 451
Score: 310.50 Matches: 87
Percent Similarity: 47.90% Conservative: 61
Best Local Similarity: 28.16% Mismatches: 146
Query Match: 19.53% Indels: 15
DB: 2 Gaps: 7

US-10-049-750-11 (1-921) x US-09-543-681A-4841 (1-451)
QY 7 ATCCGGTATTGGCTCTAATGAGTGGAGCTTATCACCTACACCAACGATGCCCAAA 66
Db 134 ValCysValLeuGlySerPheValValAspIleAlaAsnValAlaPheProArg 153
QY 67 GAAGGGGAAACTCTGAAGCGCGCGGTTTAAATCGGCTGCGCGGAAAGGGGGAAC 126
Db 154 ValGlyGluLeuValAsnSerLysGlyAsnSerIleGlyProGlyGlyLysGlyThrAsn 173
QY 127 CAGGCGCTGCGCGCGCTAAGCTCAATTCAAAAGTATTGATGTGACCAAGTGGCGGAC 186
Db 174 GlnAlaIleSerAlaSerPheSerAspAlaLysValHisLeuIleThrLysValGlyGlu 193
QY 187 GATATTTTGGCCGACACACCATTCGTAATCTCGAATCTCGGGGATCAAT-----ACG 240
Db 194 AspHisPheSerLysTyrAlaTyrLysTyrLeuGlnGluSerGlyIleAspSerPheThr 213
QY 241 ACGTATGTGAAAAGTACGCTGACAGCGCGGTAGCGCGGATTTTCGTCAAGCC 300
Db 214 IlePheGlnThrGluIleGluProThrGlySerSerIleSer---TyrLeuAlaAspLys 232
QY 301 AACTCCAGACACGATTCGTATCATCAAGGCGCTAACAAAGTTCTCTCGCGGAGAT 360
Db 233 ThrGlnAsnAsnIleThrAlaThrTyrLeuGlyAlaAsnAsnThrPheThrLeuGlnGlu 252
QY 361 ATCGATCGCGCGCGGGAAGATTTAAAAAATGCCAGCTATTGTTCTGCAACTGGAAGTT 420
Db 253 ValAspIleSerLeuProTyrIleSerGluAlaAspValLeuLeuLeuGlnGlyGluIle 272
QY 421 CAGCTTGAACGGTTTATCAGCAATAGAAATTGGCAAGAACACCGGGATTGAAGTGTAA 480
Db 273 AsnIleAspAlaAsnValLysAlaAlaSerPheAlaHisSerIleAsnLysThrValIle 292
QY 481 TTAACCTCGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAATCGGATTC 540
Db 293 LeuAsnValAlaProTyrSerAspAspLeuLysGlnLeuTyr---LeuTyrValAspPhe 311
QY 541 TTTGTACCTAATGAACACGAGCTGGAATATTAAACCGGTATGCCAGTGATACCTATGAC 600
Db 312 IleThrLeuAsnAlaTyrGlnAlaSerAspTrpSerGlyIleGluIleAsnThrIleAsn 331
QY 601 CATATTCGCGCAGCGCAGCTTCGCTGGTAGATAAAGGCTGAAACAATATTATGTCACC 660
Db 332 AspAlaLysGlnAlaValGluIleIleAlaGlyAsnGluLysLysLysValIleIleTyr 351
QY 661 ATGGCGGAGAAAGCGCGCTGTGGTAGACGCTGACGAGGAAGTCCATGTTCCGGCGTTT 720
Db 352 IleAspGluLeuGlyValValTyrPheAspGlyArgAsnThrPheHisIleProLeu 371
QY 721 AGAGTCAACGCTTGTGATACGCGCGGGGATGCTTTTATCGGCTGTTTCGGCGCAT 780
Db 372 ProSerLeuArgValAspThrMetAlaIleThrAspAlaPheAsnGlyAlaPheAlaSer 391
QY 781 TACTACGTCCAGCGGGATGTGGAAGCGCGCATGAAAGCCGCTCTCTTTGCC--- 837
Db 392 LysIleAlaAlaGlyGly-----ThrMetProGluSerValLeuPheAlaSer 407
QY 838 -----GCTTTACGCTCAGCGGGAAGGACCCCATCTCTTATCCAGCAT 885
Db 408 AlaPheLeuSerAlaPhe---IleGluGlnLysGlyVal---ThrSerMetProSerLeu 425
```

```
QY 529 AAATGCGATTCTTCTTAATGAACCGAGCTGAAATATTAAACCGGTATGCCAGTG 588
Db 43 LysThrAspIlelleProAsnGluThrGluThrGluIlelleThrGlyIleArgVal 62
QY 589 GATACCTATGACCATATTCCGCGAGCGGACCGTTCGCTGTAGATAAAGCGGTGAACAAT 648
Db 63 ThrAspHisAsnSerLeuValAlaAlaGluLysLeuHisGluLeuGlyIleGlyThr 82
QY 649 ATTATTGTCACCATGGCCAGAAAGCGCGTGTGGATGACCGGTGACCGAAGTCCAT 708
Db 83 ValIlelleThrLeuGlySerAlaGlyAlaPheTyrHisThrGluLysGluHisGlyIle 102
QY 709 GTTTCGCGCTTTAGATGAACGCTGTGTATACAGCGCGCGCGCATGCTTTATTCGCG 768
Db 103 ValProAlaPheLysValAspAlaValAspThrThrAlaAlaGlyAspThrPheIleGly 122
QY 769 TGTTCGCGCATTTACTACCTCCAGACGGGGATGTGAA-----GCCGCCATGAAA 819
Db 123 AlaLeu-----SerSerThrLeuGlnProAspLeuSerAsnLeuLys 136
QY 820 AAAGCGTCTCTTT-----GCCGCTTTCAGCGTCAACCGGGAAGCCACCCAA 867
Db 137 GluAlaIleLeuTyrGlyAsnLeuAlaSerSerValAlaValGlnSerTyrGlyAlaGln 156
QY 868 TCCTCTTATCCA-----AGCATTGAGCAA 891
Db 157 ProSerIleProTyrArgGluAlaLeuGluGln 167

RESULT 14
US-09-710-279-786
; Sequence 786, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 786
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-786

Alignment Scores:
Pred. No.: 6.47e-18 Length: 319
Score: 234.00 Matches: 81
Percent Similarity: 44.04% Conservative: 63
Best Local Similarity: 24.77% Mismatches: 137
Query Match: 14.72% Indels: 46
DB: 2 Gaps: 12

US-10-049-750-11 (1-921) x US-09-710-279-786 (1-319)
QY 16 ATTGGCTTAACATGTGTGACCTATC---ACCTACACACACAGATGCCCAAGAGGG 72
Db 7 IleGlyGluAlaLeuIleAspPheIleProAsnValThrHisSerLysLeuLysAspVal 26
QY 73 GAAACTCTGAAGCGCGGCTTTAAATCGGCTGCGCGGAAAGGGCGAACCAGGCG 132
Db 27 Glu-----GlnPheSerArgGlnValGlyGlyAlaProCysAsnValAla 41
QY 133 GTGGCGCGCCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCGGACGATTT 192
Db 42 AlaThrValSerLysLeuGlyGlyLysSerGluMetIleThrGlnLeuGlyAsnAspAla 61
```

```
QY 193 TTTGCGCAACACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACGTATGTAGAA 252
Db 62 PheGlyAspIlelleValGluThrIleGluGlnLeuGlyValGlyThrGlnTyrIleLys 81
QY 253 AAAGTACCGGTGTACAGCAGCGGTAGCGCGATTTTCGTCAACGCCCACTCCACCAAC 312
Db 82 ArgThrAsnLysAlaAsnThrAlaLeuAlaPheValSerLeuGlnAspGlyGlnArg 101
QY 313 AGCATTCCTGATCATCAAA-----GGCGCTAAACAAGTTTCTCTCGCGCAAGATATCGAT 366
Db 102 AspPheSerPheTyrArgLysProSerAlaAspMetLeuTyrGlnProGluAsnIleAsp 121
QY 367 CGCGCGCGGGAAGATTAAAAAATGCCAGCTTATTGTCTTCGCAACTGGAAGTTACGCTT 426
Db 122 -----AspIleGlnValPheGlnAspAspIleLeuHisPhe---CysSerVal 136
QY 427 GAAACGCTTTATCAGCAATAGAAATTGGCAGAAACACGGGATTCGAA-----474
Db 137 AspLeuIleGluSerAspMetLysTyrAlaHisGluLysMetIleGluLysPheGluSer 156
QY 475 -----GTGTTATTAAACCTCGCCAGCATTAACGGGAATTAGATATGCTCTTAT 522
Db 157 ValAspGlyThrIleValPheAspProAsnValArgLeuProLeuTyrGluAspLysLeu 176
QY 523 GCCTGTAAA-----TGCGATTCTTTGTACCT-----AATGAA 555
Db 177 GluCysGlnArgThrIleAsnAlaPheIleProLysAlaHisIleValLysIleSerAsp 196
QY 556 ACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATTCGCGCAGCG 615
Db 197 GluGluLeuLeuPheIleThrGlyLysArgAsnGluAsp-----GluAla 211
QY 616 GCACGTTTCGCTGTAGATAAAGCGGTGAACAATATTATTGTCCACATGCGCGAAGAGCG 675
Db 212 IleGlnSerLeuPheArgGlyGlnValAsnValIleTyrThrGlnGlyAlaGlnGly 231
QY 676 GCGCTGTGGATCAGCGTGCACAG---GAAGTCCATGTTCCGCGCTTTAGAGTGAACGCT 732
Db 232 AlaThrIleTyrThrLysAspTyrArgIleHisGluGlyTyrGlnValGlnAla 251
QY 733 GTTGATACACGCGCGCGCGATGCTTTATCGGCTGTTCGCGCATTTACTACGTCCAG 792
Db 252 IleAspThrThrGlyAlaGlyAspAlaPheIleGlyAlaIleIleTyrCysIleLeuGlu 271
QY 793 AGC-----GGGATGTGGAAGCGCCATGAAAGAACCGTCTCTC-----831
Db 272 SerArgHisSerGluCysLysAspLeuPheLysGluLysGlyLysAspIleLeuAlaPhe 291
QY 832 -----TTTGCGCGCTTTCAGCGTCAACCGGAAAGCCCAATCCTCTTATCCAAAGC 882
Db 292 SerAsnArgValAlaAlaLeuThrThrThrLysHisGlyAlaIleGluSerLeuProThr 311
QY 883 ATTGAGCAATTTAATGAGTAT 903
Db 312 LysGluAspIleLysAspTyr 318

RESULT 15
US-09-134-001C-3900
; Sequence 3900, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
```



```
; SEQ ID NO 3900
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3900

Alignment Scores:
Pred. No.:      8,48e-18      Length:      321
Score:          233.00      Matches:      82
Percent Similarity: 43.87%      Conservative: 61
Best Local Similarity: 25.15%      Mismatches: 139
Query Match:      14.65%      Indels:      44
DB:              2          Gaps:      12

US-10-049-750-11 (1-921) x US-09-134-001C-3900 (1-321)

QY 16 ATTGGCTCTAACATGTGTGACCTTATC---ACCTACCAACACAGATGCCCAAGAGGG 72
DB 9 IIEGlyGluAlaLeuileAspPheileProAsnValThrHisSerLysLeuLysAspVal 28
QY 73 GAAACTCTGGAGCGCGCGTTTAAATCGGCTCGCGCGGAAAGGGCGAACCAGGCC 132
DB 29 Glu-----GlnPheSerArgGlnValGlyGlyAlaProCysAsnValAla 43
QY 133 GTGGCGCGCCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGGACGATATT 192
DB 44 AlaThrValSerLysLeuGlyLysSerGluMetIleThrGlnLeuGlyAsnAspAla 63
QY 193 TTTGGCGGACACACCATTCGTAATCTCGAATCTCTGGGGATCAATACGAGTGTAGAA 252
DB 64 PheGlyAspIleValGluThrIleGluGlnLeuGlyValGlyThrGlnTyrIleLys 83
QY 253 AAGAGTACCGTGTACACGAGCGCGTAGCGCGGATTTTCGTCAACGCCCACTCCAGCAAC 312
DB 84 ArgThrAsnLysAlaAsnThrAlaLeuAlaPheValSerLeuGlnAspGlyGlnArg 103
QY 313 AGCATTTCTGATCATCAA-----GGCGCTAACAAAGTTTCTCTCGCGGAAGATATCGAT 366
DB 104 AspPheSerPheTyrArgLysProSerThrAspMetLeuTyrGlnProGluAsnIleAsp 123
QY 367 CGCGCGCGGAGATTAAATAATGCCAGCTTATTGTCTGCAA-----411
DB 124 -----AspIleGlnIlePheGlnAspAspIleLeuHisPheCysSerValAsp 139
QY 412 ---CTGGAAGTTCAGCTTGAAACGGTTATCAC--GCAATAGAAATTTGCCAAGAAACAC 465
DB 140 LeuileGluSerAspMetLysAsnAlaHisGluLysMetIleGluLysPheGluSerVal 159
QY 466 GGGATTGGAAGTGTATTAAACCTCGCGCCAGCATTTACGGGAATTAGATATGCTTTATGCC 525
DB 160 GlyGlyThrIleValPheAspProAsnValArgLeuProLeuTrpGluAspLysLeuGlu 179
QY 526 TGTAAG-----TGCATTTCTTTGTACCT-----AATGAAACC 558
DB 180 CysGlnArgThrIleAsnAlaPheileProLysAlaHisIleValLysIleSerAspGlu 199
QY 559 GAGCTGGAATAATTAAACCGTATGCCAGTGATACCTATGACCATATTCGCGCAGCGGCA 618
DB 200 GluLeuLeuPheIleThrGlyLysLysAsnGluAsp-----GluAlaIle 214
QY 619 CGTTCGCTGTAGATAAAGCGCTGAACAATATTATTGTCACCATGGCGGAGAAAGCGCG 678
DB 215 GlnSerLeuPheArgGlyGlnValAsnValIleTyrThrGlnGlyAlaGlnGlyAla 234
QY 679 CTGTGATGACCGTGACCCAG---GAAGTCCATGTTCCCGCGGTTTAGTGTAACGCTGTT 735
DB 235 ThrIleTyrThrLysAspAspTyrArgIleHisGluGlyTyrGlnValGlnAlaIle 254
QY 736 GATACAGCGCGCGCGGATGCTTTATCGGCTGTTTCGCGCATTTACTACGTCACAGGC 795
DB 255 AspThrThrGlyAlaGlyAspAlaPheIleGlyAlaIleTyrCysIleLeuGluSer 274
QY 796 -----GGGGATGTGGAAGCGCGCATGAAAAAAGCCGCTCCTC-----831
```

```
Db 275 ArgHisSerGluCysLysAspLeuPheLysGluLysGlyLysAspIleLeuAlaPheSer 294
QY 832 -----TTTGCCGCTTTTACGCTCACCGGGAAGGACCCCAATCTCTTTATCCAGCATT 885
DB 295 AsnArgValAlaAlaLeuThrThrThrLysHisGlyAlaIleGluSerLeuProThrLys 314
QY 886 GAGCAATTTAATGAGTAT 903
DB 315 GluAspIleLysAspTyr 320
```

Search completed: December 23, 2005, 22:44:30  
Job time : 46.391 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:42:49 ; Search time 162.356 Seconds  
(without alignments)  
4740.464 Million cell updates/sec

Title: US-10-049-750-11

Perfect score: 1590

Sequence: 1 atggatattcggtttattgg.....atcttcgttgtaacgaataa 921

Scoring table: BLOSUM62

-Q= /cgn2\_10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DBV=xlp  
-Q=/cgn2\_1/USPTO.spool.p/US10049750/runat\_23122005\_113555\_6388/app\_query.fasta\_1.1742  
-DB=Published Applications\_AA\_Main -QFMT=fastan -SUFFIX=n2p.rapbm  
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct  
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext  
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10049750 @CNG 1.1 405 @runat\_23122005\_113555\_6388 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEVI\_TIMEOUT=120 -WARN -TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA\_Main:

1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pcp:\*  
2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pcp:\*  
3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pcp:\*  
4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pcp:\*  
5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pcp:\*  
6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1519	95.5	306	4	US-10-238-075-1139
2	870	54.7	298	4	US-10-369-493-11397
3	870	54.7	298	4	US-10-369-493-14281
4	870	54.7	298	4	US-10-369-493-14829
5	870	54.7	298	4	US-10-369-493-15011
6	797	50.1	304	4	US-10-369-493-7730
7	550	34.6	308	3	US-09-815-242-11854
8	546	34.3	322	5	US-10-781-581-208
9	546	34.3	322	5	US-10-745-237-304
10	523	32.9	343	4	US-10-369-493-5584
11	507	31.9	309	4	US-10-369-493-12350

12	506	31.8	302	4	US-10-369-493-7383
13	506	31.8	303	4	US-10-369-493-4625
14	501	31.5	299	4	US-10-264-237-2526
15	499.5	31.4	300	4	US-10-369-493-7595
16	499.5	31.4	303	4	US-10-369-493-4836
17	495.5	31.2	300	4	US-10-369-493-13944
18	488.5	30.7	293	4	US-10-369-493-9151
19	476.5	30.0	304	3	US-09-815-242-13017
20	476.5	30.0	304	3	US-09-815-242-13153
21	474.5	29.8	293	4	US-10-369-493-23325
22	468.5	29.5	309	3	US-09-815-242-13883
23	468.5	29.5	318	4	US-10-369-493-22598
24	467	29.4	309	4	US-10-369-493-211
25	460	28.9	294	4	US-10-369-493-17450
26	460	28.9	304	6	US-11-097-143-31038
27	457.5	28.8	293	4	US-10-369-493-16604
28	454	28.6	303	4	US-10-369-493-20083
29	452	28.4	307	5	US-10-470-048B-270
30	452	28.4	315	4	US-10-724-972A-5080
31	450.5	28.3	306	3	US-09-815-242-11035
32	448	28.2	309	4	US-10-369-493-21162
33	445.5	28.0	309	3	US-09-741-669-479
34	445.5	28.0	309	3	US-09-815-242-10406
35	445.5	28.0	309	4	US-10-369-493-23616
36	434.5	27.3	305	3	US-09-815-242-10892
37	433	27.2	302	4	US-10-369-493-18802
38	431	27.1	302	5	US-10-501-282-3560
39	431	27.1	322	5	US-10-501-282-3562
40	415.5	26.1	340	4	US-10-369-493-4143
41	414	26.0	299	4	US-10-369-493-2985
42	412	25.9	429	5	US-10-745-237-108
43	409	25.7	285	4	US-10-369-493-8326
44	407.5	25.6	300	4	US-10-369-493-18468
45	401.5	25.3	299	4	US-10-156-761-12851

ALIGNMENTS

RESULT 1

US-10-238-075-1139  
; Sequence 1139, Application US/10238075  
; Publication No. US20030148324A1  
; GENERAL INFORMATION:  
; APPLICANT: I.N.S.E.R.M.  
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isola  
; FILE REFERENCE: BLANDINE  
; CURRENT APPLICATION NUMBER: US/10/238,075  
; CURRENT FILING DATE: 2002-09-10  
; PRIOR APPLICATION NUMBER: 0003145  
; PRIOR FILING DATE: 2000-03-10  
; NUMBER OF SEQ ID NOS: 1576  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 1139  
; LENGTH: 306  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-238-075-1139

Alignment Scores:  
Pred. No.: 1.2e-149 Length: 306  
Score: 1519.00 Matches: 294  
Percent Similarity: 98.69% Conservative: 8  
Best Local Similarity: 96.08% Mismatches: 4  
Query Match: 95.53% Indels: 0  
DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x US-10-238-075-1139 (1-306)

QY 1 ATGGATATCGGTTATTCGCTCTTAACATGGTGGACCTTATACCTACCAACCAAGATG 60  
Db 1 MetAspIleAlaValIleGlySerAsnMetValAspIleuIleThrTyThrAsnGlnMet 20

```
QY 61 CCCAAAGAGGGGAAACTCTGGAAGCGCGCGCGTATAAAATCGCGTCCGCGGAAAGGG 120
Db 21 ProLysGluGlyGluThrLeuGluAlaProAlaPheIleGlyCysGlyGlyVal 40
QY 121 GCGAAACAGCCCTGCGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTG 180
Db 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
QY 181 GCGCAGCATATTTTCCCGACACACCATTCGAATCTCGAATCCTGGGCGGATCAATAG 240
Db 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
QY 241 ACCTATGTAGAAAAGTACCGTCTACAGCAGCGCGGTAGCGCGATTTTCGTCAACGCC 300
Db 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
QY 301 AACTCCAGCAACAGCATTTCTGATCATCAAGGCGCTAAACAAGTTTCTCTCGCGGAGAT 360
Db 101 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
QY 361 ATCGATTCGCGCGCGGAGAAATTTAAAAAATGCCAGCTTATGTTCTGCAACTGGAGTT 420
Db 121 IleAspArgAlaAlaGluAspLeuLysLysCysLysLeuIleValLeuGlnLeuVal 140
QY 421 CAGCTTGAACCGCTTATCACGCAATAGAAATTTGGCAAGAAACACACGGGATTGAAGTTA 480
Db 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysGlyIleGluValLeu 160
QY 481 TTAACCCCTGCGCCAGCATTTACGGGAATTAGATATGCTTTATGCCCTGTAAATCGGATTTC 540
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180
QY 541 TTTGTACCTAATCAAAACGAGCTGGAATATTAACCGGTATGCCAGTGGATACCTATGAC 600
Db 181 PheIleProAsnGluThrGluLeuGluIleLeuThrGlyMetSerValAspThrTyrAsp 200
QY 601 CATATTCCGCGCAGCGGACGCTTCGCTGGTAGATAAAAGGCGTGAACAATATTATGTCACC 660
Db 201 HisIleArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
QY 661 ATGGCGCAGAAAGCGCGCTGTGGATGACGCGTGACAGGAAGTCCATGTTCCGCGGTTT 720
Db 221 MetSerGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
QY 721 AGAGTGAACCTGCTGATACACAGCGCGCGCGATCCCTTTATCGGCTGTTTCGCGCAT 780
Db 241 LysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 260
QY 781 TACTAGCTCCAGAGCGGGATGTGGAAAGCCGCATGAAAAAGCCGCTCTCTTTGCGGCT 840
Db 261 TyrTyrValGlnSerGlyAspValGluAlaLeuLysLysAlaAlaLeuPheAlaAla 280
QY 841 TTCAGCGTCCAGCGGAAAGCACCCCAATCCTTTATCCAGCAATTGAGCAATTAATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
QY 901 TATCTTTCGTTGAACGA 918
Db 301 PheLeuThrLeuAsnGlu 306
```

## RESULT 2

```
US-10-369-493-11397
; Sequence 11397, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
```

```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11397
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11397

Alignment Scores:
Pred. No.: 11e-81 Length: 298
Score: 870.00 Matches: 167
Percent Similarity: 76.61% Conservative: 59
Best Local Similarity: 56.61% Mismatches: 69
Query Match: 54.72% Indels: 0
DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x US-10-369-493-11397 (1-298)
```

```
QY 7 ATCGCGGTTATTGGCTCTAACATGTTGGACCTTATCACCTACACCAACAGATGCCCAA 66
Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 23
QY 67 GAAGGGGAAACTCTGGAAGCGCGCGTATAAAATCGGCTGCGCGGAAAAAGGGCGAAC 126
Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43
QY 127 CAGGCGCTGCGCGCCCTAAGCTCAATCAAAAGTATTGATGTTGACCAAGTGGCGGAC 186
Db 44 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 63
QY 187 GATATTTTTCGCGACACACCATTCGTAATCTCGAATCTGCGGATCTGCGGGATCAATACGCGTAT 246
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
QY 247 GTAGAAAAAGTACCGGTGTACCAGCAGCGCGTAGCGCCGATTTTCGTCAACGCCCACTCC 306
Db 84 ValValLysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103
QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAACAAAGTTTCTCTCGCGCGGAAGATATCGAT 366
Db 104 GluAsnSerIleLeuIleValLysGlyAlaAsnAlaAspLeuLeuProValGluValAsp 123
QY 367 CGCGCGCGGAGAGATTAAAAATCCAGCTTATTGTTCTGCAACTGGAAGTTCAGCTT 426
Db 124 LysAlaAlaAlaAspLeuLysGluCysGlyLeuIleLeuMetGlnMetGluValProVal 143
QY 427 GAAACCGTTTATCACGCAATAGAAATTTGGCAAGAAACACGGGATTGAAGTGTATTAAAC 486
Db 144 GluThrValTyrHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
QY 487 CTGCGCCCGCAGATTACGGGAATTAGATATGCTTTATGCTGTGTAATCGCGATTTCTTTGTA 546
Db 164 ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
QY 547 CCTAATGAAACCGAGCTGGAATATTAAACCGGTATGCCAGTGCATGATACCATATT 606
Db 184 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
QY 607 CGCGCAGCGCACGTTTCGCTGGTAGATAAAGGCTGCAACAATATTATTGTCACCATGGGC 666
Db 204 ValArgAlaAlaArgSerLeuIleAlaArgIleArgThrValIleValThrLeuGly 223
QY 667 GAGAAAGCGCGCTGTGGATGACGCGTGACACGGAAGTCCATGTTCCGGCGTTTAGATG 726
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
QY 727 AACGCTGTTGATACACGCGCGCGGATGCCCTTTATCGGCTGTTTCGCGCATTACTAC 786
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
```

Qy	487	CTGCGCCGACGATTTACGGGAATTAGATATGCTTTATGCTCTGTAATGCGATTTCTTTTGA	546
Db	164	ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal	183
Qy	547	CTTAATGAAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTTATGACCATATT	606
Db	184	ProAsnGluSerGluLeuLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle	203
Qy	607	CGCGCAGCGGACGCTTCGCTGGTATGATAAAGCGGTGAACAATATTATTGTACCATGGGC	666
Db	204	ValArgAlaAlaArgSerLeuIleAlaArgGlyIleArgThrValIleValThrLeuGly	223
Qy	667	GAGAAAGCGCGCTGTGGATGACCGGTGACCAAGGAGTCCATGTTCCGCGCTTTAGAGTG	726
Db	224	AlaArgGlyAlaArgMetIleThrSerAspGluIleValAlaAsnIleGluProValIlyVal	243
Qy	727	AACGCTGTTGATACAGCGCGCGCGGCGATCGCTTTATCGGCTGTTTCGCGCATTTACTAC	786
Db	244	ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr	263
Qy	787	GTCAGAGCGGGGATGTGGAAGCGGCCATGAAGAAAGCGGTCTCTTTGCCGCTTTCAGC	846
Db	264	AlaGluThrArgGluValValSerSerLeuIlyIysAlaSerLeuTyrAlaAlaHisSer	283
Qy	847	GTACCGGGGAAAGCCACCAATCTCTTTATCCAAGCATTTGAGCAA	891
Db	284	IleThrArgProGlyThrGlnIlyAlaTyrAlaSerIleAspGlu	298
RESULT 4			
US-10-369-493-14829			
; Sequence 14829, Application US/10369493			
; Publication No. US20030233675A1			
; GENERAL INFORMATION:			
; APPLICANT: Cao, Yongwei			
; APPLICANT: Hinkle, Gregory J.			
; APPLICANT: Slater, Steven C.			
; APPLICANT: Goldman, Barry S.			
; APPLICANT: Chen, Xianfeng			
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION			
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES			
; FILE REFERENCE: 38-10(52052)B			
; CURRENT APPLICATION NUMBER: US/10/369,493			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US 60/360,039			
; PRIOR FILING DATE: 2002-02-21			
; NUMBER OF SEQ ID NOS: 47374			
; SEQ ID NO 14829			
; LENGTH: 298			
; TYPE: PRT			
; ORGANISM: Agrobacterium tumefaciens			
US-10-369-493-14829			
Alignment Scores:			
Pred. No.:		1.le-81	Length: 298
Score:		870.00	Matches: 167
Percent Similarity:		76.61%	Conservative: 59
Best Local Similarity:		56.61%	Mismatches: 69
Query Match:		54.72%	Indels: 0
DB:		4	Gaps: 0
US-10-049-750-11 (1-921) x US-10-369-493-14829 (1-298)			
Qy	7	ATCGCGGTATTGGCTCTACATCGTGGACCTTATCCTACACCAACACAGATGCCAAA	66
Db	4	IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly	23
Qy	67	GAAGGGGAAACTCTGGAAGCGCGCGCGTTTAAAAATCGGCTCGCGGGAAAAAGGGCGAAC	126
Db	24	ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn	43
Qy	127	CAGCGCTGCGCGCGCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTGGCGAC	186

```
Db 44 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 63
QY 187 GATATTTTCCGACACACCATTCGTAATCTCGAATCCTGGGGATCAATACGAGTAT 246
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
QY 247 GTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGGATTTTCGTCACGCCCACTCC 306
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 ValValysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103
QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAACAAGTTTCTCTCGCGGGAAGATATCGAT 366
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 GluAsnSerIleLeuIleValysGlyAlaAsnAlaAspLeuLeuProValGluValAsp 123
QY 367 CGCGCGCGGAGAGATTTAAAAATGCGAGCTTATTTCTTCTGCAACTGGAAGTTACGTT 426
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 124 LysAlaAlaAlaAspLeuLysGluCysGlyLeuLeuMetGlnMetGluValProVal 143
QY 427 GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACGCGGATTTGAAGTGTATTAAAC 486
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 144 GluThrValTyrHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
QY 487 CTGCGCGCCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGGATTTCTTTGTA 546
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
QY 547 CCTAATGAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 184 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
QY 607 CGCGCAGCGCGATTCGCTGGTAGATAAAGGCGTGAACAATATTATTGTCACCATGGGC 666
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 ValArgAlaAlaArgSerLeuIleAlaArgIleArgThrValIleValThrLeuGly 223
QY 667 GAGAAAGCGCGCTGTGATGACGCGTGACGAGAGTCCATGTTCCGCGTTTAGAGTG 726
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
QY 727 AACGCTGTTGATACACGCGCGCGGCGATGCTTTATCGGCTGTTTCGGCGCATTTACTAC 786
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
QY 787 GTCAGAGCGGGATGTGGAAGCGCCATGAAAGAAACCGCTCTCTTTGCGGCTTTTCAGC 846
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 AlaGluThrArgGluValValSerSerLeuLysAlaSerLeuLysAlaAlaHisSer 283
QY 847 GTCACCGGAAAGCGCACCCCAATCTCTTATCCAGCATTTGAGCAA 891
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 IleThrArgProGlyThrGlnLysAlaTyrAlaSerIleAspGlu 298

RESULT 5
US-10-369-493-15011
; Sequence 15011, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15011
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15011
```

```
Alignment Scores: 11e-81 Length: 298
Pred. No.: 870.00 Matches: 167
Score: 76.61% Conservative: 59
Percent Similarity: 56.61% Mismatches: 69
Best Local Similarity: 54.72% Indels: 0
Query Match: 4 Gaps: 0
DB: 0

US-10-049-750-11 (1-921) x US-10-369-493-15011 (1-298)

QY 7 ATCCGCGTTATTGGCTCTAACATGTGTGACCTTATCACCTACACCAACGAGTGCACAA 66
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 IleGlyValValGlySerAsnMetValAspLeuIleThrTyrValAspArgMetProGly 23
QY 67 GAAGGGGAAACCTCTGAAGCGCGGCTTTAAATCGGCTGCGCGGAAAGGGCGAAC 126
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 24 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLysGlyAlaAsn 43
QY 127 CAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 44 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 63
QY 187 GATATTTTCCGACACACCATTCGTAATCTCGAATCCTGGGGATCAATACGAGTAT 246
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 83
QY 247 GTAGAAAAGTACCGTGTACACGCGGTAGCGCGGATTTTCGTCACGCCCACTCC 306
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 ValValysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 103
QY 307 AGCAACAGCATTTCTGATCATCAAGCGCTAACAAGTTTCTCTCGCGGGAAGATATCGAT 366
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 GluAsnSerIleLeuIleValysGlyAlaAsnAlaAspLeuLeuProValGluValAsp 123
QY 367 CGCGCGCGGAGAGATTTAAAAATGCGAGCTTATTTCTTCTGCAACTGGAAGTTACGTT 426
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 124 LysAlaAlaAlaAspLeuLysGluCysGlyLeuLeuMetGlnMetGluValProVal 143
QY 427 GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACGCGGATTTGAAGTGTATTAAAC 486
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 144 GluThrValTyrHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 163
QY 487 CTGCGCGCCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGGATTTCTTTGTA 546
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 183
QY 547 CCTAATGAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 184 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 203
QY 607 CGCGCAGCGCGATTCGCTGGTAGATAAAGGCGTGAACAATATTATTGTCACCATGGGC 666
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 ValArgAlaAlaArgSerLeuIleAlaArgIleArgThrValIleValThrLeuGly 223
QY 667 GAGAAAGCGCGCTGTGATGACGCGTGACGAGAGTCCATGTTCCGCGTTTAGAGTG 726
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 224 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 243
QY 727 AACGCTGTTGATACACGCGCGCGGCGATGCTTTATCGGCTGTTTCGGCGCATTTACTAC 786
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 244 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 263
QY 787 GTCAGAGCGGGATGTGGAAGCGCCATGAAAGAAACCGCTCTCTTTGCGGCTTTTCAGC 846
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 AlaGluThrArgGluValValSerSerLeuLysAlaSerLeuLysAlaAlaHisSer 283
QY 847 GTCACCGGAAAGCGCACCCCAATCTCTTATCCAGCATTTGAGCAA 891
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 IleThrArgProGlyThrGlnLysAlaTyrAlaSerIleAspGlu 298

RESULT 6
US-10-369-493-7730
```

; Sequence 7730, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 7730  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: Rhodobacter sphaeroides  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(304)  
; OTHER INFORMATION: unsure at all xaa locations  
US-10-369-493-7730

Alignment Scores:  
Pred. No.: 4,896-74 Length: 304  
Score: 797.00 Matches: 159  
Percent Similarity: 70.69% Conservative: 46  
Best Local Similarity: 54.83% Mismatches: 85  
Query Match: 50.13% Indels: 0  
DB: 4 Gaps: 0

US-10-049-750-11 (1-921) x US-10-369-493-7730 (1-304)

QY 7 ATCGCGTTATGGCTTAACATGCTGACCTTATCACTACCAACAGATGCCCAA 66  
Db 15 lleAlaValValGlySerAsnMetValAspLeuValThrTyrIleThrArgMetProAla 34  
QY 67 GAAGGGGAACTCTGGAAGCGCGGCTTTAAATCGCTCGCGGCGGAAAGGGCGAAC 126  
Db 35 ProGlyGluThrIleGluAlaProAspPheGluIleGlyCysGlyGlySerGlyAlaAsn 54  
QY 127 CAGCGCGTGGCGGCTCAAGCTCAATCAAAAGATTGATGCTGACCAAAAGTGGCGAC 186  
Db 55 GlnAlaValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 74  
QY 187 GATATTTTTCGCGACACACCACTTCGTAATCTCGAATCCTGGGGATCAATACGACGTAT 246  
Db 75 AspIlePheGlyGluAsnThrArgAsnLeuAlaGlnHisGlyIleAspIleArgHis 94  
QY 247 GPAGAAAAGTACCTGTACAGCGCGGTAGCGCGATTTTGTCAACGCCCACTCC 306  
Db 95 ValGluThrValAlaGlySerSerGlyValAlaProIlePheValGluAlaSerGly 114  
QY 307 ACCAAGATTTCTGATCATCAAGCGCTACAGTTTCTCTCGCGGAGATATCGAT 366  
Db 115 GluAsnSerIleLeuIleValGlyAlaAsnAlaAspLeuLeuProAlaAspValAsp 134  
QY 367 CGCGCGCGGAGAGATTAAATAAATGCGACGCTTTATGTTCTGCAACTGGAAGTTCAGCTT 426  
Db 135 ThrAlaGluThrLeuArgAlaAlaAspLeuIleLeuMetGlnMetGluValProArg 154  
QY 427 GAAACGGTTTATCAGCAATGAATTTGGCAGAAACACGAGGATTGAAGTGTATTAAAC 486  
Db 155 GluThrValIleHisThrValArgAlaAlaGluThrGlyValArgThrIleLeuAsn 174  
QY 487 CTGCGCGCAGCATTCAGGAAATAGATATGTTCTTATGCTGTAAATGCGATTCTTTGTA 546  
Db 175 ProAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 194  
QY 547 CCTAATGAAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606

Db 195 ProAsnGluSerGluLeuAlaLeuIleSerGlyLeuProThrGlySerGluGluIle 214  
QY 607 CGCGCAGCGGCACGTTTCGCTAGATATAAAGGCTGAACAAATATTATTGTCACCATGGC 666  
Db 215 AlaAlaAlaAlaLeuSerLeuIleGlyThrValIleValThrLeuGly 234  
QY 667 GAGAAAGCGCGCTGTGGATGACGCGTGACCAAGAGTCCATGTTCCGCGCTTTAGAGTG 726  
Db 235 GlyArgGlyAlaArgLeuValThrArgAlaGlyValValProIleAlaProValArgVal 254  
QY 727 AACGCTGTTGATACCGCGCGCGCGATGCGCTTTATCGGCTCTTTTCGCGCATTTACTAC 786  
Db 255 ThrProValAspThrThrGlyAlaGlyAspAlaPheIleGlyAlaPheAlaHisPheLeu 274  
QY 787 GTCCAGAGCGGGATGTGGAAGCGCCATGAAAGAAAGCGCTCTCTTTGCGCTTTTCAGC 846  
Db 275 AlaAlaThrGlyGluValGluGlyAlaLeuAlaHisAlaAlaArgTyrAlaAlaHisSer 294  
QY 847 GTACCGGGAAGACCGCCCAATCTCTTAT 876  
Db 295 ValThrGlyArgGlyThrGlnLysSerTyr 304

RESULT 7  
US-09-815-242-11854  
; Sequence 11854, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11854  
; LENGTH: 308  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-11854

Alignment Scores:  
Pred. No.: 3,628-48 Length: 308  
Score: 550.00 Matches: 126  
Percent Similarity: 59.33% Conservative: 52  
Best Local Similarity: 42.00% Mismatches: 122  
Query Match: 34.59% Indels: 0  
DB: 3 Gaps: 0

US-10-049-750-11 (1-921) x US-09-815-242-11854 (1-308)

```
QY 7 ATCCGGTTATTGGCTTAACATGCTGGACCTTATCACTACCAACACAGATGCCAAA 66
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 ValLeuValValGlySerLeuAsnMetAspLeuValValArgAlaProArgLeuProArg 24
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 67 GAAGGGGAAACTCTGGAAGCGCCGCGTTTAAATCGGCTCGCGGAAAGGGCGAAC 126
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
25 GlyGlyGluThrLeuAlaGlyGlnSerPheThrValProGlyGlyLeuGlyAlaAsn 44
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 127 CAGCGCTGGCGCGCTGAAGTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 GlnAlaValAlaAlaArgLeuGlyAlaGluValAlaMetIleGlyCysLeuGlyAsp 64
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 187 GATATTTTTCGCGACCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACGTAT 246
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65 AspAlaIleValGlyAspGlnLeuThrArgAlaLeuGlnAlaGluGlyIleAspCysGlnGly 84
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 247 GTAGAAAAGTACCGTGTACCAAGCGCGCTAGCGCGGATTTTCGTCACACGCCAATCC 306
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 ValGluArgValAlaGlyGluSerSerGlyValAlaLeuIleValValAspAspSer 104
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 307 AGCAACAGCATCTCATCATCAAGCGCTCAACAGTTTCTCTCGCGGAAGATATCGAT 366
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 GlnAsnAlaIleValIleValAlaGlyGlyAsnGlyHisLeuSerProAlaValLeuAla 124
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 367 CGCGCGCGGAAGATTAAATAATGCCAGCTTATTGTTCTGCAACTGGAAGTTTCAGCTT 426
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 ArgHisGluHisLeuLeuGluGlnAlaGlnValValValCysGlnLeuGluSerProLeu 144
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 427 GAAACGGTTTATCACGCAATAGAAATTTGGCAAGAAACACGGGATTGAAGTGTATTAAAC 486
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145 GluThrValGlyHisValLeuArgArgAlaHisAlaLeuGlyLysThrValIleLeuAsn 164
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 487 CCTGGCCAGCATTTACGGGAATTAGATGTTCTTATGCTGTAAATGCGATTCTTTGTA 546
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 ProAlaProAlaThrArgAspValProAlaGluTrpLeuProLeuValAspTyrLeuVal 184
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 547 CCTAATGAACCGAGCTCGAATATTAACCGTATGACCGTATGCGAGTGCATATGACCATATT 606
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 ProAsnGluThrGluSerGlnLeuLeuCysArgLeuProValAspSerLeuGluSerAla 204
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 607 CGCGCAGCGGCGATTCGCTGTAGATAAAGGGCTGAACAATATTATTGTCCACCATGGGC 666
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 GlyArgAlaAlaGluArgLeuArgGluMetGlyAlaGlyArgValIleValThrLeuGly 224
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 667 GAGAAAGCGCGCTGTGATGACCGCTGACCGAGGAGTCCATGTTCCGGCGTTTACGTG 726
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 AlaGlnGlyAlaLeuLeuValGlyGluGlyArgValGluHisPheProValAlaArgVal 244
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 727 AACGCTGTGTATACCAAGCGCGCGGCGATGCTTTATCGGCTGTTTCGCGCATTTACTAC 786
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 LysAlaLeuAspThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaAlaLeu 264
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 787 GTCCAGAGCGGGGATGTGGAAGCGCCCATGAAAGCGCGTCTCTTTTCGCGCTTTACG 846
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
265 AlaArgGlyLeuAspGluAlaAlaAlaIleArgPheGlyGlnAlaAlaAlaIleSer 284
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 847 GTCACCGGGAAGGACCAATCTCTTATTCACAGCATTTGACCAATTTAAGAGTATCTT 906
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
285 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluGluValGluArgAlaLeu 304
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 8
US-10-781-581-208
; Sequence 208, Application US/10781581
; Publication No. US20050019746A1
; GENERAL INFORMATION:
; APPLICANT: BiRx Therapeutics Ltd.
; APPLICANT: Seery, Liam
; APPLICANT: Hayes, Ian
; APPLICANT: Murphy, Finbarr
; TITLE OF INVENTION: Apoptosis-Related Kinase/GPCRs
; FILE REFERENCE: 8912/2015
; CURRENT APPLICATION NUMBER: US/10/781,581
; CURRENT FILING DATE: 2004-02-18
```

```
; PRIOR APPLICATION NUMBER: US 10/764,238
; PRIOR FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US 60/457,533
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: UK 0301566.6
; PRIOR FILING DATE: 2003-01-23
; NUMBER OF SEQ ID NOS: 226
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 208
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-781-581-208

Alignment Scores:          9.67e-48          Length:          322
Pred. No.:                546.00             Matches:         117
Score:                    58.98%             Conservative:    57
Percent Similarity:       39.66%             Mismatches:     117
Best Local Similarity:    34.34%             Indels:         4
Query Match:              5                  Gaps:          2
Db:

US-10-049-750-11 (1-921) x US-10-781-581-208 (1-322)
QY 7 ATCCGGTTATTGGCTTAACATGCTGGACCTTATCACTACCAACACAGATGCCAAA 66
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
18 ValValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProLys 37
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 67 GAAGGGGAAACTCTGGAAGCGCCGCGTTTAAATTCGGCTCGCGGAAAGGGCGAAC 126
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 ThrGlyGluThrIleHisGlyHisLysPheIleGlyPheGlyGlyLysGlyAlaAsn 57
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 127 CAGCGCTGGCGCGCTAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys 77
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 187 GATATTTTTCGCGACCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACGTAT 246
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78 AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe 97
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 247 GTAGAAAAGTACCGTGTACCAAGCGCGCTAGCGCGGATTTTCGTCACACGCCAATCC 306
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 ThrTyrGlnThrLysAspAlaAlaThrGlyThrAlaSerIleIleValAsnAsnGluGly 117
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 307 AGCAACAGCATCTCATCATCAAGGGCTCAACAGTTTCTCTCGCGGAAGATATCGAT 366
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 GlnAsnIleIleValIleValAlaGlyAlaAsnLeuLeuLeuLeuLeuLeuLeuLeu 137
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 367 CGCGCGCGGAAGATTAAATAATGCGAGCTTATTGTTCTGCAACTGGAAGTTCAGCTT 426
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 AlaAlaAlaAsnValIleSerArgAlaLysValMetValCysGlnLeuGluIleThrPro 157
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 427 GAAACGGTTTATCACGCAATAGAAATTTGGCAAGAAACACCGGATTGAAGTGTATTAAAC 486
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 AlaThrSerLeuGluAlaLeuThrMetAlaArgArgSerGlyValLysThrLeuPheAsn 177
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 487 CTGCGCGCAGCATTTACGGGAATTAGATGTTCTTATGCTGTAAATGCGATTCTTTGTA 546
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 ProAlaProAlaIleAlaAspLeuAspProGlnPheTyrThrLeuSerAspValPheCys 197
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 547 CCTAATGAACCGAGCTGGAATATTAACCGGTATGCGAGTGCATATGACCATATT 606
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 CysAsnGluSerGluAlaGluIleLeuThrGlyLeuThrValGlySerAlaAlaAspAla 217
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 607 CGCGCAGCGGACCGTTCCGCTGTAGATAAAGGGCTGAACAATATTATTGTCCACATGGGC 666
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 GlyGluAlaAlaLeuValLeuLysArgGlyCysGlnValValIleIleThrLeuGly 237
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 667 GAGAAAGC-----GCCTGTGGATGACCGGTGACCAAGGAGTCCATGTTCCGCGGTTT 720
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 AlaGluGlyCysValValLeuSerGlnThrGluProGluProLysHisIleProThrGlu 257
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 721 AGAGTGAACGCTGTTGATACCAAGCGCGCGCGATGCTTTTATCGGCTGTTTCGCGCAT 780
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```





```
Qy 67 GAAGGGAAACTCTGGAAGCGCGCGTCTTAAATCGGCTCGCGGAAAAAGGGCGCAAC 126
    |||:::|
Db 57 ProGlyGluSerValArgGlyHisAsnPhelysLeuGlySerGlyGlyValAsn 76
    |||:::|
Qy 127 CAGCGCGTGGCGCGCTAGCTCAATCAAAAGTATTGATTTGACCAAAAGTGGCGAC 186
    |||:::|
Db 77 GluAlaValAlaAlaAlaArgLeuGlyIleAsnValSerMetIleGlyMetValGlyGlu 96
    |||:::|
Qy 187 GATATTTTGGCGGACACACCATTCGTAATCTCAATCTCGAATCTCGGGGATCAATACGAGTAT 246
    |||:::|
Db 97 AspMetPheGlyAspSerAsnIleLysAspLeuSerSerAsnGlyValAspThrSerCys 116
    |||:::|
Qy 247 GTAGAAAAGTACCGTGTACCAGACGCGCGGTAGCGCGATTTTCGTCAACGCCCAACTCC 306
    |||:::|
Db 117 ValGlyArgThrLysLysThrHisThrAlaThrAlaThrIleThrValAsnLysGluGly 136
    |||:::|
Qy 307 AGCAACAGCATTCATCATCAACAAAGCGCTCAACAGTTTCTCTCGCGGAAAGATATCGAT 366
    |||:::|
Db 137 GluAsnAsnIleValValThrLeuGlyAlaAsnLeuGluMetSerProGluIleAlaAsp 156
    |||:::|
Qy 367 CGCGCGCGGAAGATTAAAAAATGCCAGCTTATTTGTTCTGCACTGGAAGTTCAGCTT 426
    |||:::|
Db 157 AlaAsnSerSerLysIleAlaGlySerLysMetValIleCysGlnGlyGluIleAlaAsp 176
    |||:::|
Qy 427 GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACCGGATTGAAGTGTATTAAAC 486
    |||:::|
Db 177 LysGlyAsnArgArgAlaPheGluIleAlaArgSerHisGlyValThrThrPheLeuAsn 196
    |||:::|
Qy 487 CTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546
    |||:::|
Db 197 ProAlaProGlyAspProAsnMetAspLysThrIleLeuGluLeuValAspIleIleCys 216
    |||:::|
Qy 547 CCTAATGAAACCGAGCTGGAATATTAAACGGTATGCCAGTGCAGTGCATATGACCATATT 606
    |||:::|
Db 217 ThrAsnGluAsnGluAlaGluPheIleThrGlyIleAlaGlnAsnAsnValLysAspAla 236
    |||:::|
Qy 607 CGCGCAGCGGCACGTTCTGCTGTAGATAAAGGCTGAACAATATTATTGTCAACATGGGC 666
    |||:::|
Db 237 GluLysAlaAlaAlaArgAlaMetLysMetGlyProGlnHisAlaIleIleThrLeuGly 256
    |||:::|
Qy 667 GAGAAAGCGCGCTGTGGATGACCGCT-----GACCAGGAAGTCCATGTTCCGCGCTT 720
    |||:::|
Db 257 GlyLysGlyValLeuLeuAlaSerLysGlyValAspAspValGluHisThrAlaValIle 276
    |||:::|
Qy 721 AGATGAAACGCTGTTGATACACGCGCGCGGCGATGCTTTATCGGCTGTTTCGCGCAT 780
    |||:::|
Db 277 LysValAspAlaValAspThrThrGlyAlaGlyAspCysPheCysGlySerLeuAlaAla 296
    |||:::|
Qy 781 TACTACGTCAGAGCGGGATGTGGAAGCCGCCATGAAAGAGCGCTCTCTTTGCCGCT 840
    |||:::|
Db 297 HisLeuValAlaGlyHisProIleSerAlaSerIleArgSerAlaAlaAsnLeuAlaAla 316
    |||:::|
Qy 841 TTCAGCGTCACCGGAAAGGACCAATCTCTTATCCAAAGCATT-----GAG 888
    |||:::|
Db 317 LeuSerValThrArgHisGlyThrGlnSerSerTyrTrpLysLeuAspGluIleArgGln 336
    |||:::|
Qy 889 CAATTTAATGATATCTT 906
    |||:::|
Db 337 GlnTyrProGluPheLeu 342
    |||:::|
```

## RESULT 11

```
US-10-369-493-12350
; Sequence 12350, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
```

```
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12350
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12350

Alignment Scores:
Pred. No.: 1.16e-43 Length: 309
Score: 507.00 Matches: 113
Percent Similarity: 57.10% Conservative: 60
Best Local Similarity: 37.29% Mismatches: 126
Query Match: 31.89% Indels: 4
DB: 4 Gaps: 4

US-10-049-750-11 (1-921) x US-10-369-493-12350 (1-309)
```

```
Qy 7 ATCCGCGTTATTGCTCTTAACATGCTGACCTTATACCTACACCAACAGATGCCCAA 66
    ::|:::|
Db 7 ValValIleLeuGlyValPheValAlaAspThrAlaTyrArgAlaAspArgGlnProArg 26
    |||:::|
Qy 67 GAAGGGAAACTCTGGAAGCGCGCGTCTTAAATCGGCTCGCGGAAAAAGGGCGCAAC 126
    |||:::|
Db 27 MetGlyGluThrIleLeuGlyAsnSerPheLysLeuGlyProGlyGlyLysGlySerAsn 46
    |||:::|
Qy 127 CAGCGCGTGGCGCGCTTAAGCTCAATTCAAAAGTATTGATTTGACCAAAAGTGGCGAC 186
    |||:::|
Db 47 GluAlaValAlaAlaGlyLysLeuGlyAlaAspThrThrPheLeuThrArgLeuGlyVal 66
    |||:::|
Qy 187 GATATTTTGGCGGACACACCATTCGTAATCTCGAATCTCGGGGATCAATACGAGTAT 246
    |||:::|
Db 67 AspAlaPheAlaAspMetAlaLysArgThrTrpGlnAspAlaGlyValLysSerAlaVal 86
    |||:::|
Qy 247 GTAGAAAAGTACCGTGTACCAGCAGCGCGTACGCGGATTTTCGTCAACGCCCAACTCC 306
    |||:::|
Db 87 IleAsp---ThrProAspSerTyrThrGlyAlaAlaTyrIlePheValGluGluGlySer 105
    |||:::|
Qy 307 AGCAACAGCATTCATCATCAAA---GGCGCTAACAAAGTTTCTCTCGCGGAAAGATATC 363
    |||:::|
Db 106 GlyAsnAsnAlaIleIleValSerProGlyAlaAlaMetLeuIleSerProThrAspIle 125
    |||:::|
Qy 364 GATCGCGCGCGGAAGATTTAAAAAATGCCAGCTTATTGTTCTGCACTGGAAGTTCAG 423
    |||:::|
Db 126 GluAlaAsnAlaAlaLeuIleArgGlyValPheValThrGlnLeuGluGlnPro 145
    |||:::|
Qy 424 CTTGAAACGGTTTATCACGCAATAGAAATTTGGCAAGAAACACACGGGATTGAAGTGTATTA 483
    |||:::|
Db 146 IleAspAlaAlaMetArgAlaLeuGluIleAlaArgGlyAlaGlyValThrThrIleLeu 165
    |||:::|
Qy 484 AACCTCTGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCCTT 543
    |||:::|
Db 166 AsnProAlaProAlaAlaLysLeuProAspArgValTyrThrLeu---CysAspTyrLeu 184
    |||:::|
Qy 544 GTACCTAATGAACCGAGCTGGAATAATTAAACGGTATGCCAGTGCATGATCTATGACCAT 603
    |||:::|
Db 185 ThrProAsnGluThrGluThrGluLeuThrGlyLeuLysValSerSerValAspGlu 204
    |||:::|
Qy 604 ATTGCGCAGCGCACGCTTCGCTGTAGATAAAGGCTGAACAATATTATTGTCCACCATG 663
    |||:::|
Db 205 AlaArgThrAlaSerGlyLysLeuGluIleLysGlyValGlyThrValIleValThrLeu 224
    |||:::|
Qy 664 GCGGAGAAAGCGCGCTGTGGATCGCTGACCGTGACCAAGGATCCATGTTTCGCGGTTTAGA 723
    |||:::|
Db 225 GlyAspLysGlyAlaLeuLeuHisSerLysGlyArgSerGluHisValProAlaIleSer 244
    |||:::|
Qy 724 GTGAAC---GCTGTTGATACACGCGCGCGGCGCATGCTTTATCGGCTGTTTCGCGCAT 780
    |||:::|
Db 245 AlaGlyProValValGluThrThrGlyAlaGlyAspAlaPheAsnGlyGlyPheAlaAla 264
    |||:::|
```



```
Db 4 ValValIleLeuGlyIleTyrValThrAspLeuThrPheArgAlaGlyArgMetProGln 23
Qy 67 GAAGGGAACTCTGGAAAGCCCGCGTGTAAATCGGCTCGCGGGAAGGGCGGAC 126
Db 24 IleGlyGluThrIleAlaGlySerAlaPheAlaMetGlyProGlyGlyLysGlySerAsn 43
Qy 127 CAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 44 GlnAlaValAlaAlaAlaArgAlaGlyAlaAspValValPheCysThrArgIleGlyAsn 63
Qy 187 GATATTTTTCGCGACAAACACCATTCGTAATCTCGAATCCTCGGGGATCAATACGACGTAT 246
Db 64 AspAlaPheGlyAlaIleAlaArgAlaThrTrpAlaAlaGluGlyIle---ThrAlaArg 82
Qy 247 GTAGAAAAGTACCGTGTACAGAGCGCGGTAGCGGATTTTCGTCAACGCCCACTCC 306
Db 83 AlaSerValIleGluGlyValSerThrGlyAlaAlaHisIlePheValAspAsnThr 102
Qy 307 AGC---AACACGATTCGATCATCAAGGCGCTAACAGTTTCTCTCGCGGAAGATATC 363
Db 103 GlyMetAsnAlaIleIleValAlaSerGlyAlaAlaGlyThrMetGluProAlaAspVal 122
Qy 364 CATCGCGCGCGAAGATTATAAAAAATGCCAGCTTATTGTCTGCAACTCGAAAGTTTCAG 423
Db 123 AspAlaIleGluAlaAspIleAlaAlaAlaArgValPheValThrGlnLeuGluGlnPro 142
Qy 424 CTTGAAACGGTTTATCAGCAATAGATTGGCAAGAAACACGGGATGAAGTGTATTA 483
Db 143 LeuGluAlaAlaArgGlyLeuGluValAlaAlaArgHisGlyValIleThrValPhe 162
Qy 484 AACCTCGCGCAGCATTCGGGAATTAGATATGCTTATGCTGCTGAAATGCGATTTCTTT 543
Db 163 AsnProAlaProAlaLeu---ProLeuAspGlyIlePheProLeuCysAspIyrIle 181
Qy 544 GTACTAATGAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGATACCTATGACCAT 603
Db 182 ThrProAsnGluThrGluAlaThrAlaLeuThrGlyValProIleAlaAsnAlaAsp 201
Qy 604 ATTCCGCGAGCGGACGTCGCTGTAGATAAAGGCTGAACAATATTATTGTCAACATG 663
Db 202 AlaArgAlaAlaAspValLeuAlaLysGlyValGlyThrValIleValThrLeu 221
Qy 664 GCGGAGAAAGCGCGCTGTGATACGCTGACGAGGAAGTCCATGTTCCGCGCTTTAGA 723
Db 222 GlyGluGlyAlaLeuLeuHisSerAlaAlaGlnSerIleLeuValProAlaIyrHis 241
Qy 724 GTG---AACCGCTGTGATACAGCGCGCGCGATGCTTTTATCGGCTGTTTCGCGCAT 780
Db 242 CysGlyArgValValGluThrAlaGlyAlaGlyAspGlyPheThrGlyGlyPheAlaAla 261
Qy 781 TACTACGTCCAGAGCGGGATGTGAAGCCGCGCATGAATAAGCGCTCTCTTTGCGCGCT 840
Db 262 AlaLeuAlaArgGlyAspAlaMetThrAlaLeuArgPheGlyCysAlaLeuAlaGly 281
Qy 841 TTCAGCTGCACGGGAAGGACCACTCTTATCCAGCATTGACGATTTAATGAC 900
Db 282 IleSerValThrArgProGlyThrAlaProSerMetProThrLeuAspGluValAsnGln 301
Qy 901 TATCTT 906
Db 302 ValLeu 303
```

RESULT 14

```
US-10-264-237-2526
; Sequence 2526, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIORITY FILING DATE: 2002-10-04
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 2526
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (213)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (215)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (277)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (282)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (285)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (296)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (297)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (299)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-2526
Alignment Scores:
Pred. No.: 4.85e-43 Length: 299
Score: 501.00 Matches: 104
Percent Similarity: 59.70% Conservative: 53
Best Local Similarity: 39.54% Mismatches: 104
Query Match: 31.51% Indels: 2
DB: 4 Gaps: 1
US-10-049-750-11 (1-921) x US-10-264-237-2526 (1-299)
Qy 7 ATCCGCGTATTGCTCTAATCATGGTGACCTTATCACCTACACCAACACGATGCCCAA 66
Db 18 ValValValValGlySerCysMetThrAspLeuValSerLeuThrSerArgLeuProLys 37
Qy 67 GAAGGGAACTCTGGAAAGCCCGCGTGTAAATCGGCTCGCGGGAAGGGCGGAC 126
Db 38 ThrGlyGluThrIleHisGlyHisPhePheIleGlyPheGlyLysGlyAlaAsn 57
Qy 127 CAGCGCTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 58 GlnCysValGlnAlaAlaArgLeuGlyAlaMetThrSerMetValCysLysValGlyLys 77
Qy 187 GATATTTTTCGCGACAAACACCATTCGTAATCTCGAATCCTCGGGGATCAATACGACGTAT 246
Db 78 AspSerPheGlyAsnAspTyrIleGluAsnLeuLysGlnAsnAspIleSerThrGluPhe 97
Qy 247 GTAGAAAAGTACCGTGTACAGAGCGCGGTAGCGGATTTTCGTCAACGCCCACTCC 306
Db 98 ThrTyrGlnThrLysAspAlaAlaThrGlyThrAlaSerIleIleValAsnAsnGluGly 117
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:44:48 ; Search time 7.5438 Seconds  
(without alignments)  
1741.646 Million cell updates/sec

Title: US-10-049-750-11  
Perfect score: 1590  
Sequence: 1 atgagatcgggtattgg.....atcttcttgtaacgaataa 921

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Fgapop 10.0, Fgapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 108002

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlp

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10049750/runat\_23122005\_113555\_6421/app\_query.fasta\_1.1742  
-DB=Published Applications AA New -QFMT=fastan -SUFFIX=n2p.rapbn -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10049750 -CGN 1 1 1 @runat\_23122005\_113555\_6421  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/PTC\_NEW\_PUB.pcp.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pcp.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	458	28.8	302	US-10-793-626-3062	Sequence 3062, Ap
2	234	14.7	319	US-10-793-626-786	Sequence 786, App
3	167.5	10.5	238	US-10-485-517-177	Sequence 177, App
4	141.5	8.9	306	US-10-485-517-185	Sequence 185, App
5	118.5	7.5	346	US-10-467-657-6514	Sequence 6514, Ap
6	101	6.4	471	US-10-467-9628-10	Sequence 10, Appl
7	100.5	6.2	4384	US-10-821-234-1120	Sequence 1120, Ap
8	91	5.7	187	US-10-793-626-448	Sequence 448, App
9	89	5.6	320	US-10-467-657-3254	Sequence 3254, Ap
10	87.5	5.5	445	US-10-858-730-209	Sequence 209, App

11	86	5.4	482	6	US-10-467-657-7660	Sequence 7660, Ap
12	85	5.3	186	6	US-10-793-626-2530	Sequence 2530, App
13	83.5	5.3	445	7	US-11-055-822-170	Sequence 170, App
14	83	5.2	244	6	US-10-467-657-3330	Sequence 3330, Ap
15	83	5.2	400	6	US-10-467-657-2702	Sequence 2702, Ap
16	82	5.2	317	6	US-10-485-517-196	Sequence 196, App
17	79.5	5.0	205	7	US-11-055-822-548	Sequence 548, App
18	79.5	5.0	205	7	US-11-055-822-1128	Sequence 1128, Ap
19	79.5	5.0	224	7	US-11-055-822-546	Sequence 546, App
20	79.5	5.0	224	7	US-11-055-822-1126	Sequence 1126, Ap
21	79.5	5.0	424	6	US-10-467-657-2678	Sequence 2678, Ap
22	79.5	5.0	525	7	US-11-074-176-146	Sequence 146, App
23	78.5	4.9	831	6	US-10-467-657-4486	Sequence 4486, Ap
24	78	4.9	233	6	US-10-793-626-416	Sequence 416, App
25	77.5	4.8	444	7	US-11-074-176-170	Sequence 170, Appl
26	77.5	4.9	445	6	US-10-873-528-30	Sequence 30, Appl
27	77.5	4.9	782	6	US-10-972-053-2	Sequence 2, Appl
28	77.5	4.9	782	6	US-10-972-053-8	Sequence 8, Appl
29	77	4.8	374	7	US-11-055-822-968	Sequence 968, App
30	77	4.7	1070	7	US-11-000-463-721	Sequence 721, App
31	77	4.7	1194	7	US-11-107-028-2	Sequence 249, Appl
32	76.5	4.8	798	7	US-11-107-028-2	Sequence 2, Appl
33	76.5	4.8	2214	7	US-11-080-991-94	Sequence 94, Appl
34	76	4.8	305	6	US-10-793-626-1362	Sequence 1362, Ap
35	76	4.8	400	6	US-10-467-657-7096	Sequence 7096, Ap
36	76	4.8	470	6	US-10-467-657-8420	Sequence 8420, Ap
37	76	4.8	625	6	US-10-793-626-2464	Sequence 2464, Ap
38	75.5	4.7	784	6	US-10-972-053-10	Sequence 10, Appl
39	75.5	4.7	792	6	US-10-972-053-12	Sequence 12, Appl
40	75	4.7	299	6	US-10-858-730-17	Sequence 17, Appl
41	75	4.7	343	6	US-10-467-657-4824	Sequence 4824, Ap
42	75	4.7	343	6	US-10-467-657-7256	Sequence 7256, Ap
43	75	4.7	348	6	US-10-467-657-4840	Sequence 4840, Ap
44	75	4.7	426	6	US-10-467-657-1548	Sequence 1548, Ap
45	75	4.7	436	6	US-10-467-657-7694	Sequence 7694, Ap

ALIGNMENTS

RESULT 1

US-10-793-626-3062  
; Sequence 3062, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; PRIORITY FILING DATE: 2004-03-04  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3062  
; LENGTH: 302  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-3062

Alignment Scores:  
Pred. No.: 5.09e-36 Length: 302  
Score: 458.00 Matches: 108  
Percent Similarity: 55.96% Conservative: 61  
Best Local Similarity: 35.76% Mismatches: 129  
Query Match: 28.81% Indels: 4  
DB: 6 Gaps: 4

US-10-049-750-11 (1-921) x US-10-793-626-3062 (1-302)

QY 7 ATCGCGGTATTGGCTCTTAACATGGTGGACCTTACCTACCAACAGATGCCAAA 66

```
Db 1 ValIleValIleGlySerThrAsnValAspLysPheLeuAsnValLysArgPheProlys 20
Qy 67 GAAGGGAACTCTGGAAGCG---CCGGCTTTAAATCGCTGGCGGGAAGGGCG 123
Db 21 ProGlyGluThrLeuHisIleAsnGlnAlaGlnLysGluPheGlyGlyGlyGlyAla 40
Qy 124 AACAGGCGCGGGCGCTCAAGCTCAATCAAAAGTATTGATGTTGACCAAGTGGC 183
Db 41 AsnGlnAlaIleAlaAlaSerArgLeuAlaAspThrPheIleSerLysValGly 60
Qy 184 GACGATATTTTGGCGCAACACCATCTGTAATCTCGAATCTCGGGGATCAATACGACG 243
Db 61 LysAspGlyAsnAlaAsnPheIleLeuGluAspPheLysAlaGlyIleHisThrGln 80
Qy 244 TATGTAGAAAAGTACCGTGTACACGAGCGGTAGCGCGATTTTCGTCACGCCAAC 303
Db 81 TyrIleLeuThrSerGluSerGluGluThrGlyGlnAlaPheIleThrValAspGluAla 100
Qy 304 TCCAGCAACAGCATCTGATCATCAAGCGCTAACAGTTTCTCTCGCGGAAGATATC 363
Db 101 GlyGlnAsnThrIleLeuValTyrGlyAlaAsnMetThrLeuSerAlaThrAspVal 120
Qy 364 GATCGCGCGGGAAGATTAAATAAAATGCCAGCTTATTGTTCTGCACTCGAAGTTGACG 423
Db 121 GluMetSerAlaAspAlaPheIleGlyAlaAspPheValValAlaGlnLeuGluValPro 140
Qy 424 CTTGAAACGGTTTATCAGCGATAGATTTCGCAAGAAACACGGGATGAAGTGTATTA 483
Db 141 PheGluAlaIleGluGlnAlaPheLysIleAlaArgLysGlnAsnIleThrThrValLeu 160
Qy 484 AACCTCGCGCAGCATTAACGGAATTAGATATGCTTATGCTGTAAATGGATTTCTTT 543
Db 161 AsnProAlaProAlaIle---GluLeuProLysSerLeuLeuLeuThrAspIleIle 179
Qy 544 GTACCTAATGAACCGAGCTGGAATATTAAACCGGTATGCCAGTGATGATACCAT 603
Db 180 IleProAsnGluThrGluAlaGluLeuLeuThrGlyIleSerIleAsnAsnGluSerAsp 199
Qy 604 ATTGCGCGAGCGCGCTGCTGATGATGACGCTGACAGGAAGTCCATGTTCCGGCGTTAGA 723
Db 200 MetLysGluThrAlaThrTyrPheLeuAspLeuGlyIleSerAlaValLeuIleThrLeu 219
Qy 664 GCGGAGAAGCGCGCTGCTGATGATGACGCTGACAGGAAGTCCATGTTCCGGCGTTAGA 723
Db 220 GlyGluGlnGlyThrTyrCysAlaTyrGlnGlnTyrLysMetIleProAlaCysAsn 239
Qy 724 GTGAACGCTGTGTATACACGCGCGCGCGATGCTTTTATCGGTGTTTC---GCGCAT 780
Db 240 ValLysAlaIleAspThrThrAlaAlaGlyAspThrPheIleGlyAlaPheLeuSerGlu 259
Qy 781 TACTAGTCCAGAGCGGGGATGTGMAAGCCGCATGAAAGAGCGTCTCTTTGCGGCT 840
Db 260 LeuAsnLysAspLeuSerAlaIleGluSerAlaIleArgLeuAlaAsnGlnAlaSerSer 279
Qy 841 TTCAGCGTCACGGGAGAGCACCAATCTTATCCACGATTTGACCAATTTAAT--- 897
Db 280 LeuThrValGlnArgLysGlyAlaGlnAlaSerIleProThrArgLysGluValGluAla 299
Qy 898 GAGTAT 903
Db 300 Glutyr 301
```

## RESULT 2

```
US-10-793-626-786
; Sequence 786, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
```

```
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 786
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-786
```

## Alignment Scores:

Pred. No.:	8.01e-15	Length:	319
Score:	234.00	Matches:	81
Percent Similarity:	44.04%	Conservative:	63
Best Local Similarity:	24.77%	Mismatches:	137
Query Match:	14.72%	Indels:	46
DB:	6	Gaps:	12

US-10-049-750-11 (1-921) x US-10-793-626-786 (1-319)

Qy	16	ATTGGCTCTAACATGGTGGACCTTATC---ACCTACACCAACACGATGCCCAAGAAGGG	72
Db	7	IleGlyGluAlaLeuIleAspPheIleProAsnValThrHisSerLysLeuLysAspVal	26
Qy	73	GAAGACTCTGGNAGCGCGCGCTTTAAATCGCTGCGCGGAAAGGGCGAACCAGGCC	132
Db	27	Glu-----GlnPheSerArgGlnValGlyGlyAlaProCysAsnValAla	41
Qy	133	GTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGACGATATT	192
Db	42	AlaThrValSerLysLeuGlyLysSerGluMetIleThrGlnLeuGlyAsnAspAla	61
Qy	193	TTTCCGCAACACCATTCGTAATCTCGAATCTCGGCGGATCAATACGACTATGTAGAA	252
Db	62	PheGlyAspIleIleValGluThrIleGluGlnLeuGlyValGlyThrGlnTyrIleLys	81
Qy	253	AAAGTACCGTGTACCAGCGGTAGCGCGGATTTTCGTCACCGCAACTCCAGCAAC	312
Db	82	ArgThrAsnLysAlaAsnThrAlaLeuAlaPheValSerLeuGlnAspAspGlyGlnArg	101
Qy	313	AGCATTTCTGATCATCAAA-----GGCGCTAACAAAGTTTCTCTCGCGGAAGATATCGAT	366
Db	102	AspPheSerPheTyrArgLysProSerAlaAspMetLeuTyrGlnProGluAsnIleAsp	121
Qy	367	CGCGCGCGGAAGATTAAATAAATGCCAGCTTATTGTTCTGCAACTGGGAAGTTCAGCTT	426
Db	122	-----AspIleGlnValPheGlnAspAspIleLeuHisPhe---CysSerVal	136
Qy	427	GAAGCGTTTATCAGCAATAGATTTTGGCAAGAAACACGGGATTGAA-----	474
Db	137	AspLeuIleGluSerAspMetLysTyrAlaHisGluLysMetIleGluLysPheGluSer	156
Qy	475	-----GTGTTATTAAACCTCGCGCACCATTTACGGGAATTAGATATGCTTTAT	522
Db	157	ValAspGlyThrIleValPheAspProAsnValArgLeuProLeuTyrGluAspLysLeu	176
Qy	523	GCCTGTAAA-----TGCAGTTCCTTGTACTT-----AATGAA	555
Db	177	GluCysGlnArgThrIleAsnAlaPheIleProLysAlaHisIleValLysIleSerAsp	196
Qy	556	ACCGAGCTGGAATATTAAACCGGTATGCCAGTGATACCTATGACCATTTCCGCGACGC	615
Db	197	GluGluLeuLeuPheIleThrGlyLysArgAsnGluAsp-----GluAla	211
Qy	616	GCAGCTTCGCTGGTAGATAAAAGGGCTGAACAATATTATTGTACCATGGCGGAGAAAGGC	675
Db	212	IleGlnSerLeuPheArgGlyGlnValAsnValIleTyrThrGlnGlyAlaGlnGly	231
Qy	676	CGCTGTGTGATGACGCGTGACCGAG---GAAGTCCATGTTCCGGCGTTTAGAGTGAACCT	732











```

Db 98 AlasertyrLeuMetArgAspLysGlnIleAlaThrThrValLysLeuArgValAla 117
Qy 265 ACCACGCGGGTACGCGCGATTTCTCAACGCGCAACTCCAGCAACAGCATCTCTGATC 324
Db 118 ArgAsnGlnIleLeuArgLeuAspPheGluGluHisProAsnArgLysValLeu--- 136
Qy 325 ATCAAGGCGCTTAAAGATTTCTCGCGGGAAGATATCGATCGCGCG---GCGGAAGAT 381
Db 137 -----GluGlnIleLysArgLysArgGluIle 146
Qy 382 TTAATAAATGCCACGCTTATTGTTCTGCAA-----CTGGAGTTTCAGCTTGAACGGTT 435
Db 147 LeuProGluTyrAspAlaIleIlePheSerAspTyrGlyLysGlyLeuSerHisIle 166
Qy 436 TATCAGCAATAGAAATTTGGCAAGAAACACACGGGATTGAAGTGTTATTAAACCTCGGCCA 495
Db 167 SerAspMetIleAspTrpAlaLysHisValGlyLysThrValLeuIleAspPro----- 184
Qy 496 GCATTACGGGAATTAGATATGCTTTATGCTGCTGTAATGCGATTTCTTTGTACCTAATGAA 555
Db 185 -----LysGlyAspTyrGluLysTyrValGlyAlaThrLeuIleThrProAsnCys 202
Qy 556 ACCGAGCTGGAATATTAAACGGTATGCCAGTGGATACCTATGACCATATTCGCCGACGG 615
Db 203 AlaGluLeuLysGluValValGlySerTrpLysAsnGluGlyAspLeuThrGluLysAla 222
Qy 616 GCACGCTCGCTGATAGATAAAGGCTGAACATATTATTGTCCACCATGGCGGCAAGAGGC 675
Db 223 GlnAsnLeuArgArgHisLeuAspLeuThrAlaValLeuLeuThrArgSerGluGluGly 242
Qy 676 GCGCTGTGGATGACGCGTACACGAGAGTCCATGTTCCGCGGTTTGGAGTGAACGCTGTT 735
Db 243 MetThrLeuPheSerGluGlyGluProIleTyrGlnProThrArgAlaGlnGluValTyr 262
Qy 736 GATACACGCGCGCGCGCATCCCTTTATC-----GGC 768
Db 263 AspValSerGlyAlaGlyAspThrValIleAlaGlyMetGlyLeuGlyLeuAlaGly 282
Qy 769 TGT 771
Db 283 Cys 283

```

## RESULT 10

```

US-10-858-730-209
; Sequence 209, Application US/10858730
; Publication No. US2005025568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgsey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 209
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum

```

## US-10-858-730-209

## Alignment Scores:

```

Pred. No.: 0.636 Length: 445
Score: 87.50 Matches: 64
Percent Similarity: 38.32% Conservative: 41
Best Local Similarity: 23.36% Mismatches: 88
Query Match: 5.50% Indels: 81
DB: 6 Gaps: 16

```

## US-10-049-750-11 (1-921) x US-10-858-730-209 (1-445)

```

Qy 85 GCGCGCGGCTTAAATCGGCTGCGGGGAAAGGGCGAACACGCGCTGCGCGCGCT 144
Db 6 AlaProSerPheAsn-----ProGlyLysGlyProGlySerAlaValGlyIleAla 22
Qy 145 AAGCTC-----AATTCAAAGTATTG-----ATGTTGACCAAGTGGCGAC 186
Db 23 LeuLeuGlyPheGlyThrValGlyThrGluValMetArgLeuMetThrGluTyrGlyAsp 42
Qy 187 GATATTTTTCGCGCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACGACGTAT 246
Db 43 GluLeu---AlaHisArgIleGlyGlyProLeuGluValArgGlyIleAlaValSerAsp 61
Qy 247 GTAGAAAAGTACCGTGTACACGAGCGCGGTAGCGCGGATTTTCGTCAACGCCCACTCC 306
Db 62 IleSerLys-----ProArgGluGlyValAlaPro----- 71
Qy 307 AGCAACAGCATTCGATCATCAAGGCGCTAACAGTTTCTCTCGCCGGAAGATATCGAT 366
Db 72 -----GluLeuLeuThrGluAspAlaPheAlaLeuIleGluArgGluAspValAsp 88
Qy 367 GCGCGCGGGAAGATTTAAATAATGCCAGCTTATTGTTCTGCAACTCGAAGTTCAGCTT 426
Db 89 IleValValGlu-----ValIleGlyGlyIleGluTyrProArg 101
Qy 427 GAACGGTTTATCACGCAATAGATTTTGGCAAGAA----- 462
Db 102 GluValValLeuAlaLeuLysAlaGlyLysSerValValThrAlaAsnLysAlaLeu 121
Qy 463 -----CACGGGATTGAAGTGTATTATAAACCTCGCGCCAGCATTCACGGAATTAGAT 513
Db 122 ValAlaAlaHisSerAlaGluLeu-----AlaAsp 131
Qy 514 ATGTCTTATGCTGTAAATGCGATTTCTTTGTACCTAATGAACACGAGCTGGAATATTA 573
Db 132 AlaAlaGluAlaAlaAsnValAspLeuTyrPhe-----GluAlaAlaValAla 147
Qy 574 ACCGTATGCGATGGATACCTATGACCATATTCCGCGAGCGGCACGCTTCGCTGTAGAT 633
Db 148 GlyAlaIleProVal-----ValGlyProLeuArgArgSerLeuAlaGly 162
Qy 634 AAAGGCTGAAACAATATTATTGTCCATGCGCGAGAAAGCGCGCTGTGGATGACGCGT 693
Db 163 AspGlnIleGlnSerValMet-----GlyIleValAsnGlyThrThr 176
Qy 694 GACCAAGGAGTCCATGTTCCGCGGTTTAGAGTGAACGCTGTTGTATACACGCGCGCG 750
Db 177 Asn-----PheIleLeuAspAlaMetAspSerThrGlyAlaAsp 189
Qy 751 ---GCGGATGCTTTTATCGGCTGTTTCGCGCATTACTACGTCAG-----ACGCGG 798
Db 190 TyrAlaAspSerLeuAlaGluAlaThrArgLeuGlyTyrAlaGluAlaAspProThrAla 209
Qy 799 GATGTGGAA-----GCCGCCATGAAAGGCGCTCCTC 831
Db 210 AspValGluGlyHisAspAlaAlaSerLysAlaAlaIleLeu 223

```

## RESULT 11

```

US-10-467-7660
; Sequence 7660, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:

```

```
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7660
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7660

Alignment Scores:
Pred. No.: 0.901 Length: 482
Score: 86.00 Matches: 75
Percent Similarity: 33.15% Conservative: 43
Best Local Similarity: 21.07% Mismatches: 118
Query Match: 5.41% Indels: 120
DB: 6 Gaps: 17

US-10-049-750-11 (1-921) x US-10-467-657-7660 (1-482)
Qy 7 ATCCGGTATTGGCTTAACATGGT-----GACCTTATC 42
Db :|||||:|||||:|||||:|||||:
10 LeuAlaValMetGlyGlnAsnLeuLeuAsnMetAsnAspCysGlyPheLysValVal 29
Qy 43 ACCTACACCAACAGATCCCAAGAGAGGGAAACTCTGGAGGCGCGCGCTTTAAATC 102
Db :|||:|||||:|||||:|||||:
30 AlaTyrAsnAArgThrIleGlyLysValAspGluPheLeuAsnGlyAlaAlaLysGluThr 49
Qy 103 GGC---TCGCGCGGNAAGGGCGACAGCCGCTGGCGCGCTTAAGCTCAATTCACAAA 159
Db :|||||:|||||:|||||:|||||:
50 GlyIleIleGlyAlaTyrSerLeuGlnAspLeuValAspLysLeuAlaLysProArgLys 69
Qy 160 GTATTGATGTTGACCAAGTGGCGAC-----GATATTTTTCGCCGACACACCATTCGT 213
Db :|||||:|||||:|||||:|||||:
70 IleMetMetValArgAlaGlySerValValAspPheValGluGlnLeuLeuPro 89
Qy 214 AATCTCGAATCCTGGGGATC-----AATACGACGTATGTAGAAAAAGTA 258
Db :|||||:|||||:|||||:|||||:
90 LeuLeuGluGluGlyAspIleIleAspGlyGlyAsnAlaAsnTyrProAspThrThr 109
Qy 259 CCGTGTACCAGC-----AGCGCGGTAGCGCGATTTTCGTCACAGCCCACTCCAGCAAC 312
Db :|||||:|||||:|||||:|||||:
110 ArgArgThrHisTyrLeuAlaGlyLysGlyIleLeuPheValGlyAlaGlyValSerGly 129
Qy 313 AGCATTCGATCATCAAGGGCGCTAACAG-----TTTCTCTCGCCCGGAAGATATCGAT 366
Db :|||||:|||||:|||||:|||||:
130 Gly-----GluGluGlyAlaArgGlyProSerIleMetProGlyGlyAspLys 146
Qy 367 CGCGCGCGGAGATTTAAAAATGCGCAGCTTATGTTCTGCACTCGGAAGTTCAG--- 423
Db :|||||:|||||:|||||:|||||:
147 ArgAlaTyrGluAlaValLysProIlePheGlnAlaIleAlaLysThrProGlnGly 166
Qy 424 -----CTTGAACCGTTTAT 438
Db :|||||:|||||:|||||:|||||:
167 GluProCysCysAspTrpValGlyLysAspGlyAlaGlyHisPheValLysMetValHis 186
Qy 439 CACCAATAGAAATTTGGCAAGAAACACGGGATTTGAAGTGTATTAAACCCCTGCGCCAGCA 498
Db :|||:|||||:|||||:|||||:
187 AsnGlyIleGluTyrGly----- 192
Qy 499 TTACGGGAATTAGATGCTTATGCTTAAATGCGATTTCTTTGTACCT----- 549
Db :|||||:|||||:|||||:|||||:
193 -----AspMetGlnLeuIleCysGluAlaTyrGlnPheMetLysAspGlyLeu 208
```

```
Qy 550 -----AATGAAACCGAGCTGAA 567
Db :|||||:|||||:|||||:|||||:
209 GlyLeuSerTyrAspGluMetHisArgValPheAlaGluTrpAsnLysThrGluLeu--- 227
Qy 568 ATATTAAACCGGTATGCCAGTGGATACCTAT---GACCATATTCCGCGAGCGGACGTTGCG 624
Db :|||||:|||||:|||||:|||||:
228 -----AspSerTyrLeuIleGluIleThrAlaAlaIleLeuGly 240
Qy 625 CTGCTAGATAAAGG-----CTGAACAATATTATTGTCACCATGGCGGAGAAA 672
Db :|||||:|||||:|||||:|||||:
241 TyrLysAspGluGlyGlyGluProLeuAlaGluLysIleLeuAspThrAlaGlyGlnLys 260
Qy 673 GCGCGCGTGTGGATGAGCGCGTGACAGGAAGTCCATGTTCCGGCGTTTAGAGTGAACGCT 732
Db :|||||:|||||:|||||:|||||:
261 GlyThrGlyLysTrpThr-----GlyIleAsnAla 270
Qy 733 GTTGAT-----ACCAGCGCGCGCGGATGCGCTTTATCGGCTGT----- 771
Db :|||||:|||||:|||||:|||||:
271 LeuAspLeuGlyIleProLeuThrLeuIleSerGluAlaValPheAlaArgCysValSer 290
Qy 772 ---TTCCGCGCATTAAGTCCAGCGCGGATGTGAAGCC----- 810
Db :|||||:|||||:|||||:|||||:
291 SerPheLysGluGlnArgValGlnThrGlyLysLeuPheAlaArgThrAlaThrProVal 310
Qy 811 -----GCCATGAAAAAGCCGCTCCTC 831
Db :|||||:|||||:|||||:|||||:
311 GluGlyGlyLysGlnGluTrpValGluAlaLeuArgGlnAlaLeuLeu 326

RESULT 12
US-10-793-626-2530
; Sequence 2530, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2530
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2530

Alignment Scores:
Pred. No.: 0.856 Length: 186
Score: 85.00 Matches: 28
Percent Similarity: 41.32% Conservative: 22
Best Local Similarity: 23.14% Mismatches: 59
Query Match: 5.35% Indels: 12
DB: 6 Gaps: 3

US-10-049-750-11 (1-921) x US-10-793-626-2530 (1-186)
Qy 535 GATTTCCTTTGTACCTAATGAACCGAGCTGGAAATATTAAACGGTATGCCAGTGGATACC 594
Db :|||||:|||||:|||||:|||||:
48 AspValValThrProAsnIleProGluAlaGluGluIleThrGlyIleLysIleAsnAsp 67
Qy 595 TATGACCATATTCGCGAGCGGCGACGTTCCGCTGGTAGATAAA---GGCTGAACATATT 651
Db :|||:|||||:|||||:|||||:
68 GluGluSerIleArgLysAlaGlyGlnIlePheIleAsnGluIleGlySerLysGlyVal 87
Qy 652 ATTGTCACCATGGCGAGAAAGCGCGCTG-----TGGATGACGCGT 693
Db :|||||:|||||:|||||:|||||:
88 ValIleLysGlyGlyHisSerAlaAspLeuAsnAlaLysAspPheLeuPheThrLys 107
```

Qy 694 GACCAGGAAGTCCATGTTCCGCGCTTTAGAGTGAACGCTGTGTGATACACGCGCGCGCGC 753  
Db 108 AsnGluThrTyThrPheGluAsnLysArgPheAspThrLysHisThrHisGlyThrGly 127  
Qy 754 GATGCTTTATCGGCTGTTCCGCGATTACTACGTCAGAGCGGGATGTGGAAGCCGCC 813  
Db 128 CysThrPheSerAlaValIleThrAlaGluLeuAlaLysGlyArgSerIleLeuAspAla 147  
Qy 814 ATGAAAAAGCGCTCTCTTTCGCCCTTTCAGCGTC-----ACCGGGAAA 858  
Db 148 ValLysLysAlaLysGluPheIleSerLeuSerIleGluHisThrProGluIleGlyLys 167  
Qy 859 GGC 861  
Db 168 Gly 168

## RESULT 13

US-11-055-822-170  
; Sequence 170, Application US/11055822  
; Publication No. US20050260707A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
; FILE REFERENCE: BGI-121CPCN  
; CURRENT APPLICATION NUMBER: US/11/055,822  
; CURRENT FILING DATE: 2005-02-11  
; PRIOR APPLICATION NUMBER: 09/606,740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141,031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142,101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148,613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187,970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19930476.9  
; PRIOR FILING DATE: 1999-07-01  
; PRIOR APPLICATION NUMBER: DE 19931415.2  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931418.7  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931419.5  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1158  
; SEQ ID NO 170  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-11-055-822-170

## Alignment Scores:

Pred. No.:	1.52	Length:	445
Score:	83.50	Matches:	61
Percent Similarity:	37.59%	Conservative:	42
Best Local Similarity:	22.26%	Mismatches:	90
Query Match:	5.25%	Indels:	81
DB:	7	Gaps:	15

US-10-049-750-11 (1-921) x US-11-055-822-170 (1-445)

Qy 85 GCGCGGCGCTTTAAATCGCTGCGCGGAAAGGGCGAACCGAGCGGTGCGCGCGCT 144  
Db 6 AlaProSerPheAsn-----ProGlyLysGlyProGlySerAlaValGlyIleAla 22

Qy 145 AAGCTC-----AATTCAAAAGTATTG---ATGTTGACCAAGTGGCGAC 186  
Db 23 LeuLeuGlyPheGlyThrValGlyThrGluValMetArgLeuMetThrGlyThrGlyAsp 42  
Qy 187 GATATTTTGGCGCAACACCATTCGTAATCTCAATCTCGGGGATCAATACGACGTAT 246  
Db 43 GluLeu---AlaHisArgIleGlyGlyProLeuGluValArgGlyIleAlaValSerAsp 61  
Qy 247 GTAGAAAAGTACCGTGTACGAGCGCGCTAGCGCGATTTTCGTCAACGCCCACTCC 306  
Db 62 IleSerLys-----ProArgGluGlyValAlaPro----- 71  
Qy 307 AGCAACGACATTCGATCATCATCAAGGGCGTAAAGTTTCTCTCGCCGGAAGATATCAT 366  
Db 72 -----GluLeuLeuThrGluAspAlaPheAlaLeuIleGluArgGluAspValAsp 88  
Qy 367 CGCGCGGGAAGATTTAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAAGTTCAGCTT 426  
Db 89 IleValValGlu-----ValIleGlyGlyIleGluTyProArg 101  
Qy 427 GAAACGGTTTATACGCAATAGAAATTTGGCAAGAAA----- 462  
Db 102 GluValValLeuAlaAlaLeuLysAlaGlyLysSerValValThrAlaAsnLysAlaLeu 121  
Qy 463 -----CACGGGATTGAAGTGTATTAAACCTCGCCGAGCATTACGGGAATTAGAT 513  
Db 122 ValAlaAlaHisSerAlaGluLeu-----AlaAsp 131  
Qy 514 ATGCTTTATGCTGTAAATGCGATTTCCTTGTACCTAATGAACCGAGCTGGAATATTA 573  
Db 132 AlaAlaGluAlaAlaAsnValAspLeuTyPhe-----GluAlaAlaValAla 147  
Qy 574 ACCGGTATGCGAGTGGATACCTATGACCATATCGCGAGCGGACGTTGCTGTGTAGAT 633  
Db 148 CysAlaIleProVal-----ValGlyProLeuArgArgSerLeuAlaGly 162  
Qy 634 AAAGGGTGAACATATTTGTCACCATGGCGGAAAGCGCGCTGTGGATGACGCT 693  
Db 163 AspGlnIleGlnSerValMet-----GlyIleValAsnGlyThrThr 176  
Qy 694 GACCAGGAAGTCCATGTTCCGCGCTTTAGAGTGAACGCTGTGTATACACGCGCGCG--- 750  
Db 177 Asn-----PheIleLeuAspAlaMetAspSerThrGlyAlaAsp 189  
Qy 751 ---GCGCATGCTTTATCGGCTGTTTCGCGCATTACTACGTCAGAGCGGGGATGTG--- 804  
Db 190 TyrAlaAspSerLeuAlaGluAlaThrArgLeuGlyTyAlaGluAlaAspProThrAla 209  
Qy 805 -----GAAGCGCCCATGAAAAAGCGGTCCTC 831  
Db 210 AsnValGluGlyHisAspAlaAlaSerLysAlaAlaIleLeu 223

## RESULT 14

US-10-467-657-3330  
; Sequence 330, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON Spa  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 3330  
; LENGTH: 244  
; TYPE: PRT

```
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3330

Alignment Scores:
Pred. No.: 1.43 Length: 244
Score: 83.00 Matches: 51
Percent Similarity: 35.47% Conservative: 21
Best Local Similarity: 25.12% Mismatches: 68
Query Match: 5.22% Indels: 63
DB: 6 Gaps: 10

US-10-049-750-11 (1-921) x US-10-467-657-3330 (1-244)
QY 363 CGATCGCGCGCGGAAGATTTAAA-----AAAATGCCAGCTTAT 401
Db 82 LysSerLysAsnGlySerPheLysThrGluLeuValSerArgSerAlaMetProArgTyr 101
QY 402 TGTTCGCAACTGGAAGTTTCAGCTTGAACCGTTTATCACGCAATAGAAATTTGCAAGAA 461
Db 102 GlnTyrThrAsnGlyArgArgIleGlnThrGly-----TipGluGlu 115
QY 462 ACA-----CGGATTGAAGTGTATTAAACCCCTGCCGACGACATT 500
Db 116 ArgAlaGluPheLysAlaGluGlyArgAspPheAspAla-LeuAsnArgPheIleAlaAs 135
QY 501 ACGGGAATTAGATATGCTTATGCTGTAATGCGATTCTTTGTACCTAATGAACCGA 560
Db 135 pValGlnThrAspAlaSerLeuGlu---AspThrAspPheSerValSerArgGluArgAr 154
QY 561 GCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATTTCGCGCAGCGGCACG 620
Db 154 GAsnGluValIleAspGlnValSerLysAspAlaValLeuArgPheLysAlaArgAlaG1 174
QY 621 TTCGCTGTGTAGAT-----AAAGGCTGAACAAATATTATTGTCACCATGGGCGGA 668
Db 174 uLysLeuAlaGlyValLeuGlyAlaSerGlyTyrLysIleValLysLeuAsnPheGlyG1 194
QY 669 GAAAGGCGCGTGTGGTAGCGGTGACCGGTGACCGAAGTCCATGTCGGGGTTTAGAGTGAA 728
Db 194 nileGlySer-----HisIle----- 199
QY 729 CGCTGTTGATACCGCGCGCGCGATGCCCTTTATCGGCTGTTTTCGGCATTTACTAGT 788
Db 200 -----AlaGlyAsp----- 202
QY 789 CCAGAGCGGGATGTGGAAGCGCGCATGAAAGCCGCTCTCTTTGCGGCTTTCAGCGT 848
Db 203 -----GlyAlaValArgAlaLysMetLeuArgAlaMetProMetAlaAla---SerVa 219
QY 849 CACCGGGAAGCGCACCAATCCTCTTATCCAGCATTGACGAATTTAATGAGTATCTTTC 908
Db 219 lAsnMetLysGlyThrAspSerAlaAlaProGlyValGluIleSer-----IleSe 237
QY 909 GTTGAAC 915
Db 237 lIleAsn 239
```

```
RESULT 15
US-10-467-657-2702
; Sequence 2702, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
```

```
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 2702
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2702

Alignment Scores:
Pred. No.: 1.64 Length: 400
Score: 83.00 Matches: 40
Percent Similarity: 36.88% Conservative: 19
Best Local Similarity: 25.00% Mismatches: 53
Query Match: 5.22% Indels: 48
DB: 6 Gaps: 8

US-10-049-750-11 (1-921) x US-10-467-657-2702 (1-400)
QY 436 TATCACGCAATAGAA-----TTTGGCAAGAAACACGGGATTGAAGTGTATTATA 483
Db 4 TyrHisAlaProAspGluLysGlyPhePheGlyGluHisGlyGlyLeuTyr-ValSerGlu 23
QY 484 AACCTCGCCGACGATTACGGGAATTAGATATGCTTAT---GCCTGTAATGCGAT--- 537
Db 24 ThrLeuIleProAlaLeuLysGluGluGlnAlaTyrAsnGluAlaLysAsnAspPro 43
QY 538 ---TTCCTTGTACCTAATGAACCGAGCTGAAATATTAAACCGGTATGCCAGTGGATACC 594
Db 44 GluPheTrpAlaGluPheArgArgAspLeuLysHisTyrValGlyArgProSerProVal 63
QY 595 TATGACCATATTTCGCGCAGCGCACGTTTCGCTGGTAGATAAAGGGCTGAACAATATTATT 654
Db 64 TyrHis-----AlaAlaArgLeuSerGluHisLeuGly----- 74
QY 655 GTCACCATGGCGGAGAAAGCGCG---CTGTGGATGACGCTGACCGAGGAAGTCCATGTT 711
Db 75 -----GlyAlaGlnIleTrpLeuLysArgGluAspLeuAsnHisThr 88
QY 712 CCGCGCTTTAGAGTGAAC----- 729
Db 89 GlyAlaHisLysValAsnAsnThrIleGlyGlnAlaLeuAlaArgArgMetGlyLys 108
QY 730 -----GCTGTTGTATACCGCGCGCGCGCGAT----- 756
Db 109 LysArgValIleAlaGluThrGlyAlaGlyGlnHisGlyValAlaSerAlaThrValAla 128
QY 757 GCCTTTATCGGCTGTTTCGCGCATTTACTGCTCAGACGGGGATGTGGAAGCCGCGCATG 816
Db 129 AlaArgPheGlyMetThrCysAspValTyrMetGlyAlaAspAspIleGlnArgGlnMet 148

Search completed: December 23, 2005, 23:20:32
Job time : 16.5438 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:20:36 ; Search time 40.015 Seconds  
(without alignments)  
4429.126 Million cell updates/sec

Title: US-10-049-750-11

Perfect score: 1590

Sequence: 1 atggatatacggttatgg.....atcttcgttgaaacgaataa 921

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10049750/runat\_23122005\_113553\_6303/app\_query.fasta\_1.1742  
-DB=PIR -QFMT=fastan -SUFFIX=en2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10049750 @CGN 1 1 83 @runat\_23122005\_113553\_6303 -NCPU=3  
-NO MMAP -LARGSEQURY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6  
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 80.\*

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1570	98.7	306	2	AE0963
2	878	55.2	312	2	AC3153
3	878	55.2	312	2	H98134
4	550	34.6	308	2	B83403
5	523	32.9	343	2	T20529
6	480.5	30.2	304	2	G89790
7	474.5	29.8	293	1	D69690
8	468.5	29.5	309	2	AB0952
9	468.5	29.5	318	2	T39594
10	463.5	29.2	308	2	AI0001
11	460	28.9	294	2	H84115
12	460	28.9	304	2	T13380
13	450.5	28.3	306	2	B64073
14	445.5	28.0	309	1	K1ECRB

15	445.5	28.0	309	2	F91215
16	445.5	28.0	309	2	G86061
17	438	27.5	307	2	AB2029
18	434.5	27.3	306	2	D82497
19	434	27.3	404	2	AF0994
20	414	26.0	299	2	E72311
21	407.5	25.6	300	2	G86829
22	399.5	25.1	313	2	AG3520
23	396	24.9	293	2	C90139
24	382.5	24.1	310	2	AF0440
25	381	24.0	300	2	A75599
26	355.5	22.4	303	2	B87367
27	352.5	22.2	378	2	F86307
28	352	22.1	299	2	D97403
29	352	22.1	299	2	AD2621
30	297.5	18.7	333	1	K1BYRB
31	291.5	18.3	305	2	T44955
32	290.5	18.3	304	2	B70680
33	288.5	18.1	300	1	S40827
34	288.5	18.1	300	2	F91229
35	288.5	18.1	300	2	E86076
36	286.5	18.0	298	2	AH0947
37	247.5	15.6	331	2	T14544
38	245	15.4	317	2	E82815
39	244	15.3	298	2	AF2933
40	244	15.3	298	2	A98349
41	241.5	15.2	325	2	B84720
42	240	15.1	310	2	G72752
43	236.5	14.9	319	2	AC0942
44	233.5	14.7	319	2	A83882
45	233.5	14.7	345	2	C86195

#### ALIGNMENTS

##### RESULT 1

AE0963  
probable carbohydrtase kinase STY3989 [imported] - Salmonella enterica subsp. enterica serov. typhi  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: This species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AB0963  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Party, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov. A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AE0963  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-306 <PAR>  
A:Cross-references: UNIPARC:UPI00000380F2; GB:AL513382; PIDN:CAD03201.1; PID:gi16504832; C:Genetics:  
C:Gene: STY3989  
C:Superfamily: ribokinase

Alignment Scores:	Length:	306
Pred. No.:	1,39e-129	306
Score:	1570.00	306
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	98.74%	Indels: 0
DB:	2	Gaps: 0

US-10-049-750-11 (1-921) x AE0963 (1-306)

QY	1	ATGGATATCGCGTTATTGGCTTAACATGGTGGACCTTATCACCTACCAACGATG 60
Db	1	MetAapIleAlaValIleGlySerAsnMetValAspLeuIleThrTyThrAsnGlnMet 20
QY	61	CCCAAGAGGGGAAACTCTGGAAGCGCGCGGCTTTAAATCGGTCGGCGGAAAGGG 120

Db	21		ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyLysGly	40
Qy	121	GCGAACACAGCGCGTGC	CGCGCTCAATTCAAAGATATTGATGTGACCAAAAGT	180
Db	41	AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal	60	
Qy	181	GGCGACGATATTTTTC	CGCAACAACCATTCGTAATCTCGAATCTCTGGGGGATCAATACG	240
Db	61	GlyAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr	80	
Qy	241	ACGTATGTAGAAAAAGTACCGTGTACCGACGCGCGGTAGCGCGGATTTTCCTCAACGCC	300	
Db	81	ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla	100	
Qy	301	AACCTCCAGCAACAGCATTTCTGATCATCAAAGCGCTAAACAAGTTTCTCTCGCGCGAAGAT	360	
Db	101	AsnSerSerAsnSerIleuIleIleLysGlyAlaAsnLysPheLeuSerProGluAsp	120	
Qy	361	ATCGATCGCGCGCGGAAGATTTAAAAAATGCCAGCTTATTGTCTGCAACTGGAAAGTT	420	
Db	121	IleAspArgAlaAlaGluAspLeuLysLysCysGlnLeuIleValLeuGlnLeuGluVal	140	
Qy	421	CAGCTTGAAACGGTTTATCAGCATAGNATTTGGCAGAAACACGGGATTCAGATGTTA	480	
Db	141	GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysHisIleGlyLeuValLeu	160	
Qy	481	TTAAACCTCGCGCCAGCATTTACGGAAATAGATATGTTATGCTCTGTAATTCGCAATTCGATTC	540	
Db	161	LeuAsnProAlaProAlaLeuArgLulLeuAspMetSerTyrAlaCysLysCysAspPhe	180	
Qy	541	TTTGTACTAATGAACCGAGCTGGAATATTAAACGGGTATCCAGTGAGTACCTATGAC	600	
Db	181	PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProValAspThrTyrAsp	200	
Qy	601	CATATTCCGCGAGCGGCAGCTTCGTGTGTAGNATAAGGCTGAACAATATTATTGTCACC	660	
Db	201	HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr	220	
Qy	661	ATGGCGAGAAAGCGCGCTGTGGATGACGCGTGACCGAAGTCCATGTTCCGCGCTTT	720	
Db	221	MetGlyGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe	240	
Qy	721	AGAGTGAACGCTGTGTATACAGCGCGCGCGGATGCTTTATCGGCTGTTTCGCGCAT	780	
Db	241	ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis	260	
Qy	781	TACTACGTCCAGCGGGGATGTGAAGCGCGCATGAAAAAGCCGTCTCTTTCGCCGT	840	
Db	261	TyrTyrValGlnSerGlyAspValGluAlaAlaMetLysLysAlaValLeuPheAlaAla	280	
Qy	841	TTACAGCTCACCGGAAAGGACCCCAATCTCTTATCCAAGCATTTGAGCAATTTAATGAG	900	
Db	281	PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu	300	
Qy	901	TATCTTTTCGTTGAACGAA	918	
Db	301	TyrLeuSerLeuAsnGlu	306	

## RESULT 2

RESUL  
AC3153

AC3153  
ribokinase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

**FIBROKINASE [IMPORTED] - AGROBACTERIUM**  
C:Species: *Agrobacterium tumefaciens*

C;Species: *Agrobacterium tumefaciens*  
C;Date: 11-Jan-2002 #sequence revision 11-Jan-2002 #text change 09-Jul-2004

C; Date: II-JAH-2002  
C: Accession: AC3153

C/ACCESSION: AC3153  
R:Wood, D.W.: Setubal, J.C.: Kaul, R.: Monks, D.: Chen, L.: Wood, G.E.: Chen, Y.: Woo, L.

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, I.; Woo, I.; Gilllet, W.; Grant, C.; Guentbner, D.; Kutvavin, T.; Lewy, R.; Li, M.; McClellan, E.

erage, G.; Gillet, W.; Grant, C.  
: Kary P.: Romero P.: Zhang S.

; Karp, P.; Romero, P.; Zhang, S.  
Science 294 2317-2323 2001

Science 294, 2317-2323, 2001  
 A. Authier, Voo H : Tao V : Bid

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon

ster, E.W.

A;Accession: AC3153  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-312 <KUR>  
A;Cross-references: UNIPROT:Q8U6G9; UNIPARC:UPI00000D2715; GB:AE008689; PIDN:AAL4564.1;  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: rbsk  
A;Map position: linear chromosome  
C;Superfamily: ribokinase

Alignment Scores:				
Pred. No.:	7.41e-69	Length:	312	
Score:	878.00	Matches:	168	
Percent Similarity:	76.25%	Conservative:	60	
Best Local Similarity:	56.19%	Mismatches:	71	
Query Match:	55.22%	Indels:	0	
DB:	2	Gaps:	0	
US-10-049-750-11 (1-921) x AC3153 (1-312)				
Qy	7	ATCGCGGTATTGGCTCTAAACATCGTGGACCTTATCACTACACCAACGATGCCCAA	66	
Db	5	IleGlyValValGlySerAsnMetValAspLeullethrIyValAspArgMetProGly	24	
Qy	67	GAAGGGGAAACTCTGGAAGCGCCGGCGTTTAAATCGGCTCGCGGGGAAAGGGCGCAAC	126	
Db	25	ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLeuGlyAlaAsn	44	
Qy	127	CAGCCGTGGCGCGCTAAAGCTCAATTCAAAGATTATGATGTTGACCAAAAGTGGCGCAC	186	
Db	45	GlnAlaValAlaAlaAArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp	64	
Qy	187	GATATTTTGGCGACAACACCAATTCGTAATCTCGAATCCTCGGGGATCAATAACGACGTAT	246	
Db	65	AspValPheAlaAspAsnThrIleAArgAsnLeuAlaSerPheGlyValAspThrArgHis	84	
Qy	247	GTGAAAAGGTACCGTGTACCAGCAGCGCGCTAGCGCGATTTTCGTCAACGCCAACATCC	306	
Db	85	ValValLysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly	104	
Qy	307	AGCAACAGCATTTCTGATCATCAAAAGGCGCTAACAAAGTTTCTCTCGCGGAAGATATCGAT	366	
Db	105	GluAsnSerIleLeulleValLysGlyAlaAsnAlaAspLeuLeuProValGluValAsp	124	
Qy	367	CGCGCGCGGAAGATTTTAAAAAATGCCAGCTTATTTCTTCGCAACTCGAAGTTCAGCTT	426	
Db	125	LysAlaAlaAlaAspLeuLysGluCysGlyLeulleLeulleLeuMetGlnMetGluValProVal	144	
Qy	427	GAACCGTTTTATCAGCAATAGATTTTGGCAAGAAACACGGGATTTGAAGTGTATTATAAC	486	
Db	145	GluThrValTyHIsThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn	164	
Qy	487	CCTCGCGCAGCATTACGGGAATAGATATGTCTTATGCTCTGTAATGCGATTTCTTTGTA	546	
Db	165	ProIleProAlaAlaAlaAsnLeuAspProGluArgIleAArgGlnValThrPheLeuVal	184	
Qy	547	CCTAATGAAACCGAGCTGGAATAATTAAACCGGTATGCCAGTGGATACCTATGACCATATT	606	
Db	185	ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle	204	
Qy	607	CGCGCAGGGCACGTTTCGTGTGTAGATAAAGGGCTGAACATATATTATTGTCCACCATGGC	666	
Db	205	ValArgAlaAlaAArgSerLeulleAlaAArgGlyIleAArgThrValIleValThrLeuGly	224	
Qy	667	GAGAAAGCGCGCTGTGTATGATCAGCGGTGACGAGGAAGTCCATGTTCGCGCGTTTAGAGTG	726	
Db	225	AlaArgGlyAlaAArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal	244	
Qy	727	AACGCTGTGATACAGCGCGCGGCGCATCCCTTTATFCGGCTGTTTCGGCATTTACTAC	786	
Db	245	ThrProArgAspThrThrGlyValGlyAspAlaPheIleGlySerPheAlaAArgPheTy	264	

165 Db ProAlaProAlaAlaAlaAsnLeuAspProGluuArgIleArgGlnValThrPheLeuVal 181

547 Qy CCTAATGAACCGAGCTGGAATATTAAACCGTATGCGAGTGATACCTATGACCAATATT 204

185 Db ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 204

607 Qy CGCGCAGCGGACGCTCGCTCGTAGATAAAGGGCTGAACAATATTATTCTCACCATGGGC 666

205 Db ValArgAlaAlaArgSerLeuIleAlaArgGlyIleArgThrValIleValThrLeuGly 224

667 Qy GAGAAAGCGCGCTGTGTGATGACGCGTGACACGAGGAAGTCCATGTTTCCGCGTTTAGAGTG 726

225 Db AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValLysVal 244

727 Qy AACGCTGTTGATACACGCGCGCGCGGCGATGCTTTATCGGCTGTTTCGCGCATTTACTAC 786

245 Db ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTyr 264

787 Qy GTCCAGACGGGGATGTGGAGCGCCCATGAAAAAGCGTCTCTTTCCCGCTTTTCAGC 846

265 Db AlaGluThrArgGluValValSerSerLeuLysLysAlaSerLeuTyrAlaAlaHisSer 284

847 Qy GTCCAGCGGAAAGGCACCAATCTCTTTATCCAAAGCATTTGACCAATTTTAATGAGTAT 903

285 Db IleThrArgProGlyThrGlnLysAlaTyrAlaSerIleAspGluPheGluAlaPhe 303

RESULT 4

B83403

ribokinase PA1950 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004

C:Accession: B83403

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, :; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportuni

A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: B83403

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-308 <STO>

A:Cross-references: UNIPROT:Q912P4; UNIPARC:UPI00000C5526; GB:AE004621; GB:AE

A:Experimental source: strain PA01

C:Genetics:

A:Gene: rbsK; PA1950

C:Superfamily: ribokinase

Alignment Scores:

Pred. No.:	4.48e-40	Length:	308
Score:	550.00	Matches:	126
Percent Similarity:	59.33%	Conservative:	52
Best Local Similarity:	42.00%	Mismatches:	122
Query Match:	34.59%	Indels:	0
DB:	2	Gaps:	0

US-10-049-750-11 (1-921) x B83403 (1-308)

Qy 7 ATCGCGGTATTGGCTCTAAACATGTGTGGAGCTTATCACCAACACAGATGCCCAAA 66

Db 5 ValLeuValValGlySerLeuAsnMetAspLeuValValArgAlaProArgLeuProArg 24

Qy 67 GAAGGGGAAACTCTCGGAAGCGCCCGGTTTAAATATCGGTGCGCGGAAAGGGCGAAC 120

Db 25 GlyGlyGluThrLeuAlaGlyGlnSerPheIleThrValProGlyGlyLysGlyAlaAsn 44

Qy 127 CAGGCGGTGGCGCGCTAAGCTCAATTCAAAGTATTGATTGACCAAGTGGCGAC 180

Db 45 GlnAlaValAlaAlaArgLeuGlyAlaGluValAlaMetIleGlyCysLeuGlyAsp 64

Qy 187 GATATTTTGGCGCAACACCATTCGTAAATCTCGAATCTCTGGGGGATCAATACGCGTAT 240

Db 65 AspAlaTyrGlyAspGlnLeuTyrArgAlaLeuGlnAlaGluGlyIleAspCysGlnGly 84

```
Qy 247 GTAGAAAAGTACCGTGTACCAGCAGCGCGTAGCGCGATTTTCGTCAACGCCAATCC 306
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 ValGluArgValAlaGlyGluSerSerGlyValAlaLeuIleValValAspSerSer 104
Qy 307 AGCAACAGCATCTGATCATCAAAAGGCGCTAACAGTTTCTCTCGCGGAAGATATCGAT 366
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 GlnAsnAlaIleValIleValAlaGlyGlyAsnGlyHisLeuSerProAlaValLeuAla 124
Qy 367 CGCGCGGCGAAGATTAAAAAATGCCAGCTATTGTTCTGCACTGGAAGTTCAGCTT 426
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 ArgHisGluHisLeuLeuGluGlnAlaGlnValValValCysGlnLeuGluSerProLeu 144
Qy 427 GAAACGGTTTATACGCAATAGATTTCGGCAAGAAACACAGCGGATTGGAAGTCTTATAAC 486
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 145 GluThrValGlyHisValLeuArgArgAlaHisAlaLeuGlyLysThrValIleLeuAsn 164
Qy 487 CTCTGCGCAGCATTACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 ProAlaProAlaThrArgAspValProAlaGluTrpLeuProLeuValAspTyrLeuVal 184
Qy 547 CTTAATGAAACCGAGCTGGAATATTACCGGTATGCCAGTATGCCAGTATACCATATT 606
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 ProAsnGluThrGluSerGluLeuLeuCysArgLeuProValAspSerLeuGluSerAla 204
Qy 607 CGCGCAGCGCAGCTTCGCTGCTAGATAAAGCGCTGAACAATATTATTGTCACCATGGC 666
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 205 GlyArgAlaAlaGluArgLeuValGlyGluGlyArgValGluHisPheProValAlaArgVal 224
Qy 667 GAGAAAGCGCGCTGTGGATGACGCGTACAGCAAGTCCATGTTCCGCGCTTTAGAGTG 726
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 225 AlaGlnGlyAlaLeuLeuValGlyGluGlyArgValGluHisPheProValAlaArgVal 244
Qy 727 AACCGTTGTATACAGCGCGCGCGCGATCCCTTTATCGCTGTTTCGCGCATTACTAC 786
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 LysAlaLeuAspThrThrAlaAlaGlyAspThrPheValGlyGlyPheAlaAlaLeu 264
Qy 787 GTCCAGAGCGGGATGTGAAGCGCGCATGAAAAAGCGCTCTTTGCGCGCTTTCAGC 846
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 265 AlaArgGlyLeuAspGluAlaAlaAlaIleArgPheGlyGlnAlaAlaAlaIleSer 284
Qy 847 GTCCCGGGAAGCACCACCAATCTCTTATCCAAAGCATTTGACCAATTTAATGATATCTT 906
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 285 ValThrArgLeuGlyAlaGlnThrSerIleProSerArgGluValGluValArgAlaLeu 304

RESULT 5
T20529
hypothetical protein F07A11.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T20529
R;Palmer, S.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19287
A;Accession: T20529
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-343 <WIL>
A;Cross-references: UNIPROT:Q19133; UNIPARC:UPI000008223A; EMBL:Z66511; PIDN:CAA91318.1;
A;Experimental source: clone F07A11
C;Genetics:
A;Gene: CESP:F07A11.5
A;Map position: 2
A;Introns: 19/1; 49/2; 136/1; 221/3; 259/1; 284/3
C;Superfamily: ribokinase
```

```
Alignment Scores:
Pred. No.: 1,07e-37 Length: 343
Score: 523.00 Matches: 114
Percent Similarity: 56.54% Conservative: 59
Best Local Similarity: 37.25% Mismatches: 127
Query Match: 32.89% Indels: 6
DB: 2 Gaps: 2
```

```
US-10-049-750-11 (1-921) x T20529 (1-343)
Qy 7 ATCCGGCTTATTGCTCTTAACATCGTGGACCTTATCATTACACCAACCATGCCCAA 66
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 37 IleValValPheGlySerIleValGlnAspLeuValSerTyrThrAspSerPheProArg 56
Qy 67 GAAGGGAAATCTCTGGAAGCGCGCGTTTAAAAATCGCTCGCGCGGAAAAAGGCGCGAAC 126
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 ProGlyGluSerValArgGlyHisAsnPhenylsLeuGlySerGlyGlyLysGlyAlaAsn 76
Qy 127 CAGCGCTGGCGCGCGCTAAAGCTCAATTCAAAAAGTATTGATTGTTCACCAAAAGTGGCGAC 186
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 77 GlnAlaValAlaAlaAlaArgLeuGlyLeuAsnValSerMetIleGlyMetValGlyGlu 96
Qy 187 GATATTTTTCGCCGACACACCATTCGTAATCTCAAAAGTATTGATTGTTCACCAAAAGTGGCGAC 246
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 97 AspMetPheGlyAspSerAsnIleLysAspLeuSerSerAsnGlyValAspThrSerCys 116
Qy 247 GTAGAAAAGTACCGTGTACCAGCAGCGCGTAGCGCGATTTTCGTCAACGCCAATCC 306
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 117 ValGlyArgThrLysLysThrHisThrAlaThrAlaThrIleThrValAsnLysGluGly 136
Qy 307 AGCAACAGCATCTGATCATCAAAAGGCGCTAACAGTTTCTCTCGCGGAAGATATCGAT 366
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 137 GluAsnAsnIleValValThrLeuGlyAlaAsnLeuGluMetSerProGluIleAlaAsp 156
Qy 367 CGCGCGCGGAGATTAAAAAATGCCAGCTTATTGTTCTGCACTGGAAGTTCAGCTT 426
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 157 AlaAsnSerSerIleAlaGlySerLysMetValIleCysGlnGlyGluIleAspGlu 176
Qy 427 GAAACGGTTTATACGCAATAGAAATTTGGCAAGAAACACAGCGGATTGAAAGTGTATTAAAC 486
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 177 LysGlyAsnArgArgAlaPheGluIleAlaArgSerHisGlyValThrPheLeuAsn 196
Qy 487 CTCTGCGCAGCATTACCGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 ProAlaProGlyAspProAsnMetAspLysThrIleLeuGluLeuValAspIleCys 216
Qy 547 CTTAATGAAACCGAGCTGGAATATTACCGGTATGCCAGTATGCCAGTATCTATGACCATATT 606
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 217 ThrAsnGluAsnGluAlaGluPheIleThrGlyIleAlaGlnAsnAsnValLysAspAla 236
Qy 607 CGCGCAGCGCAGCTTCGCTGCTAGATAAAGCGCTGAACAATATTATTGTCACCATGGC 666
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 237 GluLysAlaAlaArgAlaMetLeuLysMetGlyProGlnHisAlaIleIleThrLeuGly 256
Qy 667 GAGAAAGCGCGCTGTGGATGACCGCT-----GACCAGGAAGTCCATGTTCCGCGCGTTT 720
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 257 GlyLysGlyValLeuLeuAlaSerLysGlyValAspAspValGluHisThrAlaValIle 276
Qy 721 AGAGTGAACGCTGTTGATACAGCGCGCGCGCGATGCCCTTTATCGCTGTTTCGCGCAT 780
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 277 LysValAspAlaValAspThrThrGlyAlaGlyAspCysPheCysGlySerLeuAlaAla 296
Qy 781 TACTACGTCCAGAGCGCGGATGTGAAGCGCGCATGAAAAAGCGCTCTCTTTGCGCGCT 840
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 297 HisLeuValAlaGlyHisProIleSerAlaSerIleArgSerAlaAlaAsnLeuAlaAla 316
Qy 841 TTCAGCGCTCACCGGGAAGGCAACCAATCTCTTATCCAAAGCATTT-----GAG 888
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 317 LeuSerValThrArgHisGlyThrGlnSerSerTyrTrpLysLeuAspGluIleArgGln 336
Qy 889 CAATTTAATGATATCTT 906
Dy |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 337 GlnTyrProGluPheLeu 342

RESULT 6
G89790
hypothetical protein rbsK [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: G89790
```

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A>Title: Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: G89790

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-304 <KUR>

A:Cross-references: UNIPROT:Q99WV7; UNIPARC:UPI00000CA992; GB:BA000018; PID:gl3700184; E

A:Experimental source: strain N315

C:Genetics:

A:Gene: rbsK

C:Superfamily: ribokinase

Alignment Scores:  
Pred. No.: 5.6e-34 Length: 304  
Score: 480.50 Matches: 105  
Percent Similarity: 55.45% Conservative: 63  
Best Local Similarity: 34.65% Mismatches: 126  
Query Match: 30.22% Indels: 9  
DB: 2 Gaps: 4

US-10-049-750-11 (1-921) x G89790 (1-304)

Qy 7 ATCGGGTTATTGGCTTAACATGTCGGAGCTTATCACCTACCAACACAGATGCCCAA 66  
Db 5 ValValIleLeuGlySerThrAsnValAspGlnPheLeuThrValGluArgTyrAlaGln 24  
Qy 67 GAAGGGGAACCTCTGGAAGCG---CCGGCGTTTAAATCGGTCGGCGGAAAGGGCG 123  
Db 25 ProGlyGluThrLeuHisValGluAlaGlnLysAlaPheGlyGlyLysGlyAla 44  
Qy 124 AACAGCGCGTGGCGCGCTTAAGCTCAATCAAAAGTATTGATTTTACCAAGTGGC 183  
Db 45 AsnGlnAlaIleAlaThrAlaArgMetGlnAlaAspThrPheIleThrLysIleGly 64  
Qy 184 GACGATATTTTGGCGGACACACCATTCGTATCTCGAATCTGGGATCAATACGACG 243  
Db 65 ThrAspGlyValAlaAspPheIleLeuGluAspPheLysAlaAlaHisIleAspThrSer 84  
Qy 244 TATGTAGAAAAGTACCGTGTACACGACGCGGTAGCCGCGATTTTCGTCAACGCCAAC 303  
Db 85 TyrIleIleLysThrThrGluAlaLysThrGlyGlnAlaPheIleThrValAsnAlaGlu 104  
Qy 304 TCCAGCAACAGATTCTGATCATCAAGCGGTAAAGTTTCTCTCCGCGGAAGATATC 363  
Db 105 GlyGlnAsnThrIleTyrValTyrGlyGlyAlaAsnMetThrMetThrProGluAspVal 124  
Qy 364 GATCGCGCGCGGAAGATTAAAAAATGCCAGCTTATTGTCTGCAACTGGAAGTTCAG 423  
Db 125 IleAsnAlaLysAspAlaIleIleAsnAlaAspPheValValAlaGlnLeuGluValPro 144  
Qy 424 CTTGAAACGGTTTATCAGCAATAGAAATTTGGCAAGAAACACGGGATGAAGTGTATTA 483  
Db 145 IleProAlaIleIleSerAlaPheGluIleAlaLysAlaHisGlyValThrThrValLeu 164  
Qy 484 AACCTCGCGCCAGCATTAAGGAAATTAGATATGCTTATGCTGCTGAATATGCAATTTCT 543  
Db 165 AsnProAlaProAla---LysAlaLeuProAsnGluLeuLeuSerLeuIleAspIleIle 183  
Qy 544 GTACCTATGAACCGAGCTGGAATATTAATTAACCGGTATGCCAGTGGATACCTGACCAT 603  
Db 184 ValProAsnGluThrGluAlaGluLeuSerGlyIleLysValThrAsnGluGlnSer 203  
Qy 604 ATTCTCGCGCGCGCGCTTCGCTGTAGATAAAGGCTGCAACATATATTATTGTCACCATG 663  
Db 204 MetLysAspAsnAlaAsnTyrPheLeuSerLeuGlyIleLysThrValLeuIleThrLeu 223  
Qy 664 GCGGAGAAGCGCGCTGTGGATGACGCGTACCGAGTACAGGAAGTCCATGTTCCGCGGTTAGA 723  
Db 224 GlyLysGlnGlyThrTyrPheAlaThrLysAsnGlnSerGlnHisIleGluAlaTyrLys 243

Qy 724 GTGAACGCTGTTGTATACACAGCGCGCGGCATCCCTTTATCGGCTGTTTCGCGCATTAC 783  
Db 244 ValAsnAlaIleAspThrThrAlaAlaGlyAspThrPheIleGlyAlaPheValSerArg 263  
Qy 784 TACGTCCAGACGCGGGATGTGAAGCCCGCATGAAAAAGCGCTCTCTTT----- 834  
Db 264 LeuAsnLysSerGlnAsp-----AsnLeuAlaAspAlaIleAspPheGlyAsnLys 280  
Qy 835 ---GCCGTTTCACGCGTCACCGGAAAGGACCCCAATCTCTTATCCAGCATGAGCAA 891  
Db 281 AlaSerSerLeuThrValGlnLysHisGlyAlaGlnAlaSerIleProLeuLeuGluGlu 300  
Qy 892 TTTAATGAG 900  
Db 301 ValAsnGln 303

#### RESULT 7

D69690 ribokinase (EC 2.7.1.15) - *Bacillus subtilis*

C:Species: *Bacillus subtilis*

C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004

C:Accession: D69690; I40463; S42711

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle

teck, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.

koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidis, A.; Lardinois,

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Levine, A.; Liu, H.; Masuda, S.; Maue

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrak, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A. *Bacillus subtilis*.

A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: D69690

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-293 <KUN>

A:Cross-references: UNIPROT:P36945; UNIPARC:UPI000006084E; GB:Z99122; GB:AL009126; NID:5

A:Experimental source: strain 168

R:Woodson, K.; Devine, K.M.

Microbiology 140, 1829-1838, 1994

A>Title: Analysis of a ribose transport operon from *Bacillus subtilis*.

A:Reference number: I40462; MUID:95005437; PMID:7921236

A:Accession: I40463

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-283 'DKK' <RES>

A:Cross-references: UNIPARC:UPI0000168934; EMBL:Z25798; NID:9397493; PIDN:CAA81049.1; PI

C:Genetics:

A:Gene: rbsK

C:Superfamily: ribokinase

C:Keywords: phosphotransferase; ribose metabolism

Alignment Scores:  
Pred. No.: 1.87e-33 Length: 293  
Score: 474.50 Matches: 114  
Percent Similarity: 55.30% Conservative: 53  
Best Local Similarity: 37.75% Mismatches: 124  
Query Match: 29.84% Indels: 11  
DB: 1 Gaps: 4

US-10-049-750-11 (1-921) x D69690 (1-293)

Qy 4 GATATCGGGTTATTGGCTTAACATGTTGGACCTTATCACCTACCAACACGATGCC 63

Db 3 AsnIleCysValIleGlySerCysSerMetAspLeuValThrSerAspLysArgPro 22





```
Qy 127 CAGCGCGTGGCGCGCTTAAGCTCAATTCAAAAGTATTGATGTTGACCAAAAGTGGCGAC 186
Db ||||||||||| : : : : : |||||||
46 GlnAlaValAlaGlyArgSerGlyAlaAspIleAlaPheIleAlaCysValGlyAsp 65
Qy 187 GATATTTTGGCGGCAACACCATTCGTATCTCGAATCCTGGGGATCAATACGAGTAT 246
Db ||| : : : : : ||| : : :
66 AspAspIleGlyGluArgValArgGlnGlnLeuThrAlaAspLysIleAspThrGlnPro 85
Qy 247 GTAGAAAAGTACCGTGTACCAGCGCGGTAGCGGATTTTCGTCAACGCCCAACTCC 306
Db ||| : : : : : ||| : : : : :
86 IleGluAlaIleLysGlyAlaThrThrGlyValAlaLeuIlePheValAsnSerAspGly 105
Qy 307 AGCAACAGCATTCGATCATCAAGAGCGCTAACAGTTTCTCTCGCGGAGAGATCGAT 366
Db ||| ||| ||| ||| ||| ||| : : : : :
106 GluAsnValIleGlyIleAsnAlaGlyAlaAsnSerAlaValThrProGluTyLeuArg 125
Qy 367 CGCGCGGGAGAGATTAAAAAATGCGCAGCTTATTGTTCTGCAACTGGAAGTTTCAGCTT 426
Db ||| : : : : : ||| : : : : :
126 ArgTyrglnGlnGlnValIleAspAlaAspAlaLeuLeuMetGlnLeuGluSerProLeu 145
Qy 427 GAAACGGTTTATCAGCAGTATGATTTGGCAAGAAACACGGGATTTGAAGTCTTATTAAC 486
Db : : : ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
146 AspThrValIleAlaAlaAlaLysLeuAlaLysGlnHisGlnThrGlnValIleLeuAsn 165
Qy 487 CTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTCTGTAATGCGATTTCTTTGTA 546
Db ||||||||| ||| : : : ||| |||
166 ProAlaProAla--ArgLysLeuProAspGluLeuLeuThrLeuValAspMetIleThr 184
Qy 547 CCTAATGAAACCGAGCTGGAAATATTAAACGGTATGCCAGTGGATACCTATGACCATATT 606
Db ||||||||| ||| ||| ||| ||| : : : : :
185 ProAsnGluThrGluAlaGluArgLeuThrGlyIleHisIleGluGlnAspAspAla 204
Qy 607 CGCGCAGCGGACGTTCTCGTGTAGATAAAGGCTGAACATATATTGTCACCATGGC 666
Db ||| ||| ||| ||| ||| ||| : : : ||| : : : ||| : : : |||
205 AlalysAlaAlaGlnIleHisAspLysGlyIleAlaThrValIleIleThrLeuGly 224
Qy 667 GAGAAAGCGCGCTGTGATGACGCGTGACAGGAAGTCCAT--GTTCCGCGCTTTAGA 723
Db : : : ||| : : : ||| : : : ||| ||| ||| : : :
225 SerArgGly--ValTrpLeuSerGluGlnGlyLysGlyLysLeuValAlaGlyPheLys 243
Qy 724 GTGAACGCTGTGTATACAGCGCGCGCGGATGCTTTTATCGGCTGTTTCGCGGATTAC 783
Db ||||||||| ||| ||| ||| ||| ||| |||
244 ValAsnAlaValAspThrIleAlaAlaGlyAspThrPheAsnGlyAlaLeuLeuThrAla 263
Qy 784 TACGTCCAGCGGGGATGTGAAGCCCATGAAAGGCGCTCTTTTCGCGCTTC 843
Db : : : : : ||| : : : ||| ||| ||| |||
264 LeuLeuGluGlyGlnSerMetAspValAlaValArgPheAlaHisAlaAlaAlaIle 283
Qy 844 AGCGTCACCGGAAAGCAGCCCAATCCCTCTTATCCAAGCATTGAGCAATTTAATGAGTAT 903
Db : : : ||| ||| ||| ||| ||| : : :
284 AlaValThrArgProGlyAlaGlnProSerIleProTrpArgAlaGluIleAspSerPhe 303
Qy 904 CTT 906
Db |||
304 Leu 304

RESULT 11
H84115
ribokinasase rbsK [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: H84115
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: H84115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <STO>
A:Cross-references: UNIPROT:Q9KG61; UNIPARC:UPI00001333E4; GB:AP001519; GB:BA000004; NID
A:Experimental source: strain C-125
C:Genetics:
```

```
A:Gene: rbsK
C:Superfamily: ribokinasase

Alignment Scores:
Pred. No.: 3.5e-32 Length: 294
Score: 460.00 Matches: 113
Percent Similarity: 55.81% Conservative: 55
Best Local Similarity: 37.54% Mismatches: 117
Query Match: 28.93% Indels: 16
DB: 2 Gaps: 5

US-10-049-750-11 (1-921) x H84115 (1-294)
Qy 4 GATATCGCGGTTATTGGCTCTTAACATGGTGGAGCTTATCACCTACACCAACGAGTGGCC 63
Db : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
5 AsnIleThrValValGlySerIleAsnMetAspMetValThrIleThrAspValValPro 24
Qy 64 AAGAAGGGGAACTCTGGAAGCGCGCGGTTTAAATCGGCTCGCGGGAAGGGGCG 123
Db : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
25 ValGlnGlyGluThrValLeuGlyLysAspPheArgThrValProGlyGlyLysGlyAla 44
Qy 124 AACGAGCGCGTGGCGCGCTTAAGCTCAATTCAAAAGTATTGATGTTGACCAAAAGTGGC 183
Db ||||||||| ||| : : : ||| : : : ||| : : : ||| : : : |||
45 AsnGlnAlaValAlaAlaAlaArgLeuGlyAlaAsnValArgMetIleGlyArgValGly 64
Qy 184 GACGATATTTTTCGCGCAACACCATTCGTAATCTCGAATCCTGGGGATCAATACGACG 243
Db ||||||||| ||| ||| ||| ||| ||| |||
65 AspAspProPheGlyHisValLeuThrGluAsnLeuAlaLysGluGlyIleIleThrAsp 84
Qy 244 TATGTAGAAAAGTACCGCTGTACAGCAGCGCGGTAGCGGATTTTCGTCACCGCCAC 303
Db : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
85 SerValLysProValThrAspCysThrSerGlyValAlaThrIleLeuLeu--SerAsp 103
Qy 304 TCCAGCACGACATTCGATCATCAAGGGCGCTAACAGTTTCTCTCGCGGAAGATATC 363
Db ||| ||| ||| ||| ||| ||| ||| : : : ||| : : : ||| : : : |||
104 ArgAspAsnArgIleIleValThrLysGlyAlaAsnGluHisValThrProAspTrpVal 123
Qy 364 GATCGCGCGGGAAGATTTAAAAAATGCGCAGCTTATTGTTCTGCAACTGGAAGTTTCAG 423
Db : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
124 AlaAlaPheGluGlnGluLeuAlaAlaSerAspValValLeuLeuGlnLeuGluIlePro 143
Qy 424 CTTGAAACGCTTTATACGCAATAGAAATTTGGCAAGAAACACGGGATTTGAAGTGTATTA 483
Db ||| : : : ||| : : : ||| ||| ||| ||| ||| : : : |||
144 LeuGluThrValAlaIleValLeuGluPheCysAlaLysHisValThrValLeu 163
Qy 484 AACCTCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTT 543
Db ||| ||| ||| ||| ||| ||| ||| : : : ||| : : : ||| : : : |||
164 AsnProAlaProAlaGlnLysLeuProAspAlaAlaIleTrpThr--AspAlaThrTyrlle 182
Qy 544 GTACCTAATGAAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCAT 603
Db : : : ||| ||| ||| ||| ||| ||| ||| : : : ||| : : : |||
183 SerProAsnGluAsnGluCysLeuGlnLeuPheGlyAspGluProAspAla-----Asn 200
Qy 604 ATTCGCGCAGCGCACGTTTCGCTGCTAGATAAAGGCTGAACAATATTATTGTACACCATG 663
Db : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
201 LeuArg-----GlnLysLeuIleMetThrLys 209
Qy 664 GCGGAGAAAGCGCGCTGTGGATGACGCGTGACAGGAAGTCCATGTTCCGCGCTTTAGA 723
Db ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
210 GlyAlaAspGlyValGlnPheTyrgluAsnAspGluGlnValGlnValGlnUserPheArg 229
Qy 724 GTGAACGCTGTGTATACAGCGCGCGCGGATGCTTTTATCGGCTGTTTCGCGCATTTAC 783
Db ||| ||| ||| ||| ||| ||| ||| : : : ||| : : : ||| : : : |||
230 ValGluProValAspThrThrGlyAlaGlyAspThrPheAsnGlyAlaPheAla---Val 248
Qy 784 TACGTCCAGACGGGGATGTGAAGCGCGCATGAAAGGCGCTCTCTTTTCGCGCTTC 843
Db : : : ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
249 AlaLeuGlyGlyThrValLysGluAlaValArgPheAlaAsnAlaAlaAlaLeu 268
Qy 844 AGCGTCACCGGAAAGCAGCCCAATCTCTTATCCAAGCATTGAGCAATTTAATGAGTAT 903
Db ||||||||| ||| ||| ||| ||| ||| : : : ||| : : : ||| : : : |||
269 SerValGlnSerPheGlyAlaGlnGlyMetProThrLysAlaGlnValGlnSerPhe 288
```



```
Qy 904 CTT 906
Db 289 Leu 289

RESULT 12
Ti3380
ribokinase homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:115C2.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C:Accession: T13380
R:Salles, C.; Valenti, P.; Darlameitou, A.; Henderson, N.; Campbell, L.; Glover, D.
submitted to the EMBL Data Library, May 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17665
A:Accession: T13380
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-304 <CAT>
A:Cross-references: UNIPROT:O77425; UNIPARC:UPI000008263F; EMBL:AL031581; PIDN:CAA20884.
C:Genetics:
A:Map position: FlyBase:FBgn0020381
A:Introns: 18/2; 201/2
A:Note: EG:115C2.1
C:Superfamily: ribokinase

Alignment Scores:
Pred. No.: 3,53e-32 Length: 304
Score: 460.00 Matches: 109
Percent Similarity: 56.48% Conservative: 61
Best Local Similarity: 36.21% Mismatches: 119
Query Match: 28.93% Indels: 12
Gaps: 6
DB:

US-10-049-750-11 (1-921) x T13380 (1-304)

Qy 4 GATATCGCGTATTGGTCTTAACATGGTGGACCTTATCACCTACCAACACAGATGCC 63
Db 5 GluValLeuValPheGlySerAlaIleleAaspPheIleSerTyThrArgLeuPro 24

Qy 64 AAAGAGGGGAAACTCTGGAAGCGCGGTTTAAATCGGTCGGCGGCGGAAAGGGCG 123
Db 25 LysAlaGlyGluThrLeuHisGlyHisArgPheGlnIleGlyTyThrGlyLysGlyAla 44

Qy 124 AACAGCGCGTGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGGC 183
Db 45 AsnGlnCysValAlaAlaAlaArgGlnGlySerArgThrAlaLeuValAlaLysLeuGly 64

Qy 184 GACGATATTTTGGCGACACACCATTCGTAACTCTCGAATCTGGGGATCAATACGACG 243
Db 65 AlaAspThrPheGlySerAspTyrlleuArgHisLeuArgGluGluArgValAsnValAsn 84

Qy 244 TATGTAGAAAAGTACCGGTGTACACGACGCGCGTAGCGCGATTTTGTCAACGCCAAC 303
Db 85 HisValGluGlnLeuAlaGluThrGlyValAlaGlnIleAlaValSerAspGly 104

Qy 304 TCACGACAGCATTCGTGATCATCAAGCGCGTCAAGTTTCTCTCGCGGGAAGATATC 363
Db 105 GlyGluAsnAsnIleIleValValGlyAlaAsnAsnArgLeuSerSerCysAspVal 124

Qy 364 GATCGCGCGGGAAGATTTAAATAATGCCAGCTTATTGTTCTGCAACTGGAAGTTCAG 423
Db 125 SerSerAlaLysAlaLeuPheGlnGluAlaLysValLeuValCysGlnLeuGluThrPro 144

Qy 424 CTTGAAACGGTTTATCACGCAATAGAAATTTGGCAAGAAACACGGGATGAAGTGTATTA 483
Db 145 ValGluAlaThrLeuThrAlaLeuArgAlaPheArg-----GlyValSerile---Val 161

Qy 484 AACCTCGCCAGCATTCGGGAATTAGATATGCTTATGCTGCTGTAATGCAATTCCTTT 543
Db 162 AsnAlaAlaProAlaMetAlaAspThrProGluLeuLeuGlnLeuAlaSerilePhe 181
```

```
Qy 544 GTACCTAATGAACCGAGCTGGAATATTAAACCGGTATGCCA---GTGGATACCTATGAC 600
Db 182 CysValAsnGluSerGluAlaAlaLeuMetThrGlnMetProAspIleGlyAsnIleGlu 201

Qy 601 CATATTCGCGCAGCGCACGCTTCGCTGTGTAGATAAAGGCTGAACAATATTATTGTACC 660
Db 202 HisAlaGluAspAlaValGlyLysLeuIleAlaAlaGlyAlaAsnThrValIleIleThr 221

Qy 661 ATGGCGGAGAAAGCGCGCTGTGGATGACCGGTGACACAGGAA-----GTCCATGTT 711
Db 222 LeuGlyLysLeuGlyAlaValPheGlySerAlaAspSerLysGlyValCysGlnHisVal 241

Qy 712 CCGCGGTTAGAGTG-----AACGCTGTGTATACCGCGCGCGGCGATGCCCTTT 762
Db 242 AlaAlaProSerValProGluLysValValAspThrThrGlyAlaGlyAspAlaPhe 261

Qy 763 ATCGGCTGTTTCGCGCATTTACTAGCTCCAG-----AGCGGGGATGTGGAGCCGCCATG 816
Db 262 IleGlyAlaLeuAlaHisAsnLeuAlaArgHisProThrArgLysLeuGluGluHisIle 281

Qy 817 AAAAAAGCGTCTCTTTGCGGCTTTCAGCTCCCGGGAAGGACCAATCTCTTAT 876
Db 282 AlaAlaCysAlaValAlaSerGlnSerValGlnLeuProGlyThrGlnSerSerPhe 301

Qy 877 CCA 879
Db 302 Pro 302

RESULT 13
B64073
ribokinase (EC 2.7.1.15) - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C:Accession: B64073
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.W.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64073
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-306 <TIGR>
A:Cross-references: UNIPROT:P44331; UNIPARC:UPI00001333E5; GB:U32732; GB:L42023; NID:915
C:Superfamily: ribokinase
C:Keywords: phosphotransferase; ribose metabolism

Alignment Scores:
Pred. No.: 2,41e-31 Length: 306
Score: 450.50 Matches: 108
Percent Similarity: 53.33% Conservative: 52
Best Local Similarity: 36.00% Mismatches: 139
Query Match: 28.33% Indels: 1
Gaps: 2
DB:

US-10-049-750-11 (1-921) x B64073 (1-306)

Qy 7 ATCGCGTATTGGTCTTAACATGGTGGACCTTATCACCTACCAACACAGATGCCAAA 66
Db 5 LeuThrValLeuGlySerIleAsnAlaAspHisValIleSerValProTyThrLys 24

Qy 67 GAAGGGGAAACTCTCGAAGCGCGGCTTTAAATCTGGCTGCGGGGAAAGGGCGAAC 126
Db 25 ProGlyGluThrLeuThrGlyGlnAsnTyrlleAlaTyThrGlyLysGlyAlaAsn 44

Qy 127 CAGGCGTGGCGCGCTCAAGCTCAATCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 45 GlnAlaValAlaAlaAlaArgLeuGlyAlaLysValAlaPheIleSerCysIleGlySer 64

Qy 187 GATATTTTGGCGACACACCATTCGTAATCTCGAATCTCGGATCTCGGGGATCAATACGACGTAT 246
```





THIS PAGE BLANK (USE)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:15:36 ; Search time 248.618 Seconds  
(without alignments)  
5227.239 Million cell updates/sec

Title: US-10-049-750-11  
Perfect score: 1590  
Sequence: 1 atggtatcggtattgg.....atcttctgtgaacgaataa 921

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10049750/runat\_23122005\_113552\_6292/app\_query.fasta\_1.1742  
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HRAPIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10049750@cgn\_1.1.614 @runat\_23122005\_113552\_6292 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAIDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80:.\*  
1: uniprot\_sprot:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1570	98.7	306	Q7CPF5_SALTY	Q7cpf5 salmonella
2	1570	98.7	306	Q8XGW5_SALTI	Q8xgw5 salmonella
3	1570	98.7	306	Q5PN81_SALPA	Q5pn81 salmonella
4	1563	98.3	306	Q57143_SALCH	Q57143 salmonella
5	1519	95.5	306	Q67FX7_ECOLI	Q67fx7 escherichia
6	1519	95.5	306	Q8FKR4_ECOLI	Q8fkr4 escherichia
7	1519	95.5	315	Q70712_ECOLI	Q70712 escherichia
8	1435.5	90.3	314	Q9A119_ECOLI	Q9a119 escherichia
9	1137.5	71.5	309	Q7MP36_VIBVY	Q7mp36 vibrio vuln
10	928	58.4	309	Q6ALS2_DSPPS	Q6als2 desulfotale
11	924	58.1	320	Q4LSV5_9BURK	Q4lev5 burkholderi
12	900	56.6	308	Q4ZMV5_PSESY	Q4zmv5 pseudomonas
13	892	56.1	310	Q63M06_BURPS	Q63m06 burkholderi
14	878	55.2	312	Q8U6G0_AGR75	Q8u6g0 agrobacteri
15	831.5	52.3	304	Q6A8F4_PROAC	Q6a8f4 propionibac
16	773.5	48.6	307	Q88Z61_LACPL	Q88z61 lactobacill

17	581.5	36.6	308	2	Q896T9_CLOTE	Q896t9 clostridium
18	550	34.6	308	2	Q912F4_PSEAE	Q912f4 pseudomonas
19	549	34.5	323	2	Q8R1Q9_MOUSE	Q8r1q9 mus musculus
20	546	34.3	322	1	RBSK_HUMAN	Q9h477 homo sapien
21	539	33.9	311	2	Q61QD5_BRARE	Q61qd5 brachydanio
22	533	33.5	310	2	Q5M7S1_XENTR	Q5m7s1 xenopus tro
23	523	32.9	343	2	Q19133_CAEEL	Q19133 caenorhabdi
24	518	32.6	306	2	Q88316_PSESM	Q88316 pseudomonas
25	518	32.6	312	2	Q620B1_CAEBR	Q620b1 caenorhabdi
26	517	32.5	306	2	Q4ZUH4_PSESY	Q4zuh4 pseudomonas
27	512.5	32.2	321	2	Q8A401_BACTN	Q8a401 bacteroides
28	507	31.9	309	2	Q982U3_RHILO	Q982u3 rhizobium 1
29	501	31.5	302	2	Q88K34_PSEPK	Q88k34 pseudomonas
30	494.5	31.1	314	2	Q7UUS5_RHOBA	Q7uus5 rhodospirell
31	493.5	31.0	306	2	Q8RD45_THETN	Q8rd45 thermoanaer
32	490.5	30.8	302	2	Q9X4M5_LACSK	Q9x4m5 lactobacill
33	488.5	30.7	309	2	Q891L9_CLOTE	Q891l9 clostridium
34	487.5	30.7	305	2	Q4KEX0_PSEF5	Q4kex0 pseudomonas
35	486.5	30.6	308	2	Q4LUI3_9BURK	Q4lui3 burkholderi
36	486.5	30.6	308	2	Q6DB84_ERWCT	Q6db84 erwinia car
37	480.5	30.2	304	2	Q6GK43_STAAR	Q6gk43 staphylococ
38	480.5	30.2	304	2	Q7A7T7_STAAN	Q7a7t7 staphylococ
39	480.5	30.2	304	2	Q99WV7_STAAM	Q99wv7 staphylococ
40	478.5	30.1	304	2	Q6GCK4_STAAS	Q6gck4 staphylococ
41	478.5	30.1	304	2	Q8NYG7_STAAN	Q8nyg7 staphylococ
42	478.5	30.1	307	2	Q5FLF7_LACAC	Q5flf7 lactobacill
43	477.5	30.0	311	2	Q4LRB7_9BURK	Q4lrb7 burkholderi
44	476.5	30.0	297	2	Q5KUX1_GEOKA	Q5kux1 geobacillus
45	476.5	30.0	304	2	Q5HJA8_STAAC	Q5hja8 staphylococ

#### ALIGNMENTS

RESULT 1

Q7CPF5\_SALTY Q7CPF5\_SALTY PRELIMINARY; PRT; 306 AA.

AC Q7CPF5\_SALTY PRELIMINARY; PRT; 306 AA.  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Putative sugar kinase (SC 2.7.1.15).  
GN OrderedLocusNames=STM3793;  
OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.

RC STRAIN=LT2;

RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2."

RL Nature 413:852-856(2001).

DR EMBL; AE008876; AAL22651.1; -: Genomic DNA.

DR GO; GO:00046301; F:kinase activity; IEA.

DR GO; GO:0004747; F:ribokinase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR GO; GO:0006014; P:D-ribose metabolism; IEA.

DR InterPro; IPRO11877; D:ribokin\_bact.

DR InterPro; IPRO11811; PfkB region.

DR InterPro; IPRO02139; Ribokinase.

DR Pfam; PF00294; PfkB; 1.

DR PRINTS; PR00990; RIBOKINASE.

DR TIGRFAMs; TIGR02152; D\_ribokin\_bact; 1.

KW Complete proteome.

SEQUENCE 306 AA; 33229 MW; 142A0289E2D1B953 CRC64;

## Alignment Scores:

```
Pred. No.: 4.34e-129 Length: 306
Score: 1570.00 Matches: 306
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.74% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q7CPF5_SALTY (1-306)
Qy 1 ATGGATATCGCGGTTATTGGCTCTAACATGCTGACCTTATCACCTACCAACAGATG 60
Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyThrAsnGlnMet 20
Qy 61 CCCAAGAGGGGAACTCTGAGAGCCGCGCTTTAAATCGGCTGGCGGAAAGGG 120
Db 21 ProlysGluGlyGluThrLeuGluAlaProAlaPheIleGlyCysGlyGlyGly 40
Qy 121 GCGAACACGAGCGTGGCGCGCTAAAGTCAATCAAAAGTATTGATGTTCCACCAAGTG 180
Db 41 AlaGlnAlaValAlaAlaAlaIleGlyLeuAsnSerIleValLeuMetLeuThrIleVal 60
Qy 181 GCGACGATATTTTGGCCGACACACCATTCGTAATCTCGAATCTCGGGGATCAATACG 240
Db 61 GlyAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
Qy 241 ACGTATGTAGAAAAGTACCGTGTACACGAGCGCGTAGCGCGATTTTCGTCAACGCC 300
Db 81 ThrTyValGluIleValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
Qy 301 AACTCCAGCAACGATTCATCATCAAGCGCGCTAACAAAGTTCCTCTCGCCGGAAGAT 360
Db 101 AsnSerSerAsnSerIleLeuIleIleGlyAlaIleValPheLeuSerProGluAsp 120
Qy 361 ATCGATCGCGCGGAGAGATTTAAAAAATGCCAGCTTATTGTCTGCAATCGAAGTT 420
Db 121 IleAspArgAlaAlaGluAspLeuIleValGlnLeuGluVal 140
Qy 421 CAGCTTGAACGGTTTATCAGCAATAGAAATTTGCGAAGAACACGGGATGAGTCTTA 480
Db 141 GlnLeuGluThrValTyHisAlaIleGluPheGlyLysIleValGluVal 160
Qy 481 TTAACCTCGCGCAGCATTAACGGAATTAGATATGCTTATGCTGTAAATCGCATTC 540
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyAlaCysLysCysAspPhe 180
Qy 541 TTTGTACCTAATGAACACGAGCTGGAATATTAACCGGTATGCCAGTGATACCTATGAC 600
Db 181 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProValAspThrTyAsp 200
Qy 601 CATATTCCGCGAGCGGACGTTCCGTGTAGATAAAGGCTGAACATATTATTGTACCC 660
Db 201 HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleValThr 220
Qy 661 ATGGCGAGAAGGCGCGCTGTGATGACGCTGACACGAGGAGTCCATGTTCCGCGGTTT 720
Db 221 MetGlyGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
Qy 721 AGAGTGAACGCTGTGTATACAGCGCGCGCGGATGCTTTATCGGCTGTTTCGCGAT 780
Db 241 ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260
Qy 781 TACTACGTCCAGCGGGGATGTGGAAGCCGCCATGAAGAGCGCTCTCTTTCGCGCT 840
Db 261 TyrTyValGlnSerGlyAspValGluAlaAlaMetLysLysAlaValLeuPheAlaAla 280
Qy 841 TTCAGCGTCACCGGAAAGGACCAATCTCTTATCCAAGCATTTGACCAATTAATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyProSerIleGluGlnPheAsnGlu 300
Qy 901 TATCTTTTCGTTCAACGAA 918
Db 301 TyrLeuSerLeuAsnGlu 306
```

```
RESULT 2
Q8XGW5_SALTY
ID Q8XGW5_SALTY PRELIMINARY; PRT; 306 AA.
AC Q8XGW5; Q7AL29;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 10-MAY-2005 (TReMBLrel. 30, Last annotation update)
DE Putative carbohydrate kinase
GN OrderedLocusNames=STY3989, t3725;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyonanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Church C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AE016846; AA071218.1; -; Genomic DNA.
DR EMBL; AL627280; CAD03201.1; -; Genomic DNA.
DR HSSP; P05054; 1RKD.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D:ribokin_bact.
DR InterPro; IPR011611; PfkB region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfkB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D:ribokin_bact; 1.
KW Complete proteome; Kinase.
SQ SEQUENCE 306 AA; 33229 MW; 142A0289E2D1B953 CRC64;
```

## Alignment Scores:

```
Pred. No.: 4.34e-129 Length: 306
Score: 1570.00 Matches: 306
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.74% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q8XGW5_SALTY (1-306)
Qy 1 ATGGATATCGCGGTTATTGGCTCTAACATGCTGACCTTATCACCTACCAACAGATG 60
Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyThrAsnGlnMet 20
Qy 61 CCCAAGAGGGGAACTCTGAGAGCCGCGCTTTAAATCGGCTGGCGGAAAGGG 120
Db 21 ProlysGluGlyGluThrLeuGluAlaProAlaPheIleGlyCysGlyGlyGly 40
```



```
Db 181 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProValAspThrTyrAsp 200
Qy 601 CATATTCGCGCAGCGCGCGCTCGTGTAGATAAAGCGCTGAACAATATTATTGTACCC 660
Db 201 HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
Qy 661 ATGGCGGAGAAGCGCGCTGTGGATGACGGTGCACAGGAGTCCATGTTCCGCGCTTT 720
Db 221 MetGlyGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
Qy 721 AGAGTGAACGCTGTTGTATPACCAAGCGCGCGCGATGCCTTTATCGGCTGTTTCGCGCAT 780
Db 241 ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260
Qy 781 TACTACGTCCAGAGCGGGGATGTGAAGCCGCCCATGAAAAAGCGCTCTTTGCCGCT 840
Db 261 TyrTyrValGlnSerGlyAspValGluAlaAlaMetLysLysAlaValLeuPheAlaAla 280
Qy 841 TTCACGCTCACCGGAAAGGCACCAATCCTTATCAAGCATTTGAGCAATTTAATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
Qy 901 TATCTTTCTGTTGAACGAA 918
Db 301 TyrLeuSerLeuAsnGlu 306

RESULT 4
Q57143 SALCH PRELIMINARY; PRT; 306 AA.
ID Q57143
AC Q57143
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Putative sugar kinase, ribokinase family.
GN Name=rbsK; OrderedLocusNames=SC3713;
OS Salmonella cholerae-suis (Salmonella enterica).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SC-B67;
RX PubMed=15781495;
RA Chiu C.-H., Fang P., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,
RA Wang H.-S., Lee Y.-S.;
RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a
RT highly invasive and resistant zoonotic pathogen.";
RL Nucleic Acids Res. 33:1690-1698(2005).
DR EMBL; AE017220; AAX67619.1; -; Genomic_DNA.
KW Complete proteome; Kinase.
SQ SEQUENCE 306 AA; 33187 MW; 159A0289E2D1A942 CRC64;

Alignment Scores:
Pred. No.: 1.79e-128 Length: 306
Score: 1563.00 Matches: 305
Percent Similarity: 99.67% Conservative: 0
Best Local Similarity: 99.67% Mismatches: 1
Query Match: 98.30% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q57143_SALCH (1-306)
Qy 1 ATGGATATCGCGTTATTTGGTCTTAACATGTTGACCTTATCACCAACCAAGATG 60
Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyrThrAsnGlnMet 20
Qy 61 CCCAAGAGCGGAACTCTGAGAGCGCGCGTTTAAATCGGCTGGCGGGAAGGG 120
Db 21 ProlysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyLysGly 40
Qy 121 GCGAACCAAGCGCTGGCGCGCGCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTG 180
```

```
Db 41 AlaAsnGlnAlaValAlaAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60
Qy 181 GGCCACCATATTTTTCGCCGCAACACCATTCGTAAATCTCGAATCTCTGGGATCAATACG 240
Db 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTrpGlyIleAsnThr 80
Qy 241 ACGTATGTAGAAAAAGTACCGTGTACCAGCAGCGCGTAGCGCGATTTTCGTCAACGCC 300
Db 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100
Qy 301 AACTCCAGCAACACCATTCGATCATCAAGGCCCTAACAGTTTCTCTCCCGGAGAT 360
Db 101 AsnSerSerAsnSerIleLeuIleIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120
Qy 361 ATCCATCGCGCGCGGAGATTAAAAAATGCCAGCTTATTGTTCTGCACTCGAAGTT 420
Db 121 IleAspArgAlaAlaGluAspLeuLysCysGlnLeuIleValLeuGlnLeuGluVal 140
Qy 421 CAGCTTGAACCGTTTATCAGCAATAGAAATTTGCGCAAGAAACACGGGATTGAAGTGA 480
Db 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysHisGlyIleGluValLeu 160
Qy 481 TTAACCCCTCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGTAAATGCGATTTC 540
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180
Qy 541 TTTGTACTTAATGAACCGAGCTCGAAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600
Db 181 PheValProAsnGluThrGluLeuGluIleLeuThrGlyMetProValAspThrTyrAsp 200
Qy 601 CATATTCGCGCAGCGCGCATCGTTCGTGTAGATAAAGCGCTGAACAATATTATTGTACCC 660
Db 201 HisIleArgAlaAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220
Qy 661 ATGGCGGAGAAGCGCGCTGTGTGATGACGGTGCACAGGAGTCCATGTTCCGCGCTTT 720
Db 221 MetGlyGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240
Qy 721 AGAGTGAACGCTGTTGATACCGCGCGCGCGATGCTTTATCGGCTGTTTCGCGCAT 780
Db 241 ArgValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaHis 260
Qy 781 TACTACGTCCAGAGCGGGGATGTGAAGCCGCCCATGAAAAAGCGCTCTCTTTGCCGCT 840
Db 261 TyrTyrValGlnSerGlyAspValGluAlaAlaMetLysLysAlaGlyLeuPheAlaAla 280
Qy 841 TTCAGCTCACCGGAAAGGCACCAATCCTTATCAAGCATTTGAGCAATTTAATGAG 900
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300
Qy 901 TATCTTTCTGTTGAACGAA 918
Db 301 TyrLeuSerLeuAsnGlu 306

RESULT 5
Q67FX7 ECOLI PRELIMINARY; PRT; 306 AA.
ID Q67FX7_ECOLI
AC Q67FX7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Deoxyribokinase.
GN Name=dek;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=55989, AL862, and EC185;
RX PubMed=15385522; DOI=10.1128/IAI.72.10.6151-6156.2004;
RA Bernier-Februe C., Du Merle L., Turlin E., Labas V., Ordenez J.,
RA Gilles A.M., Le Bouguenec C.;
```



"Use of Deoxyribose by Intestinal and Extraintestinal Pathogenic  
Escherichia coli Strains: a Metabolic Adaptation Involved in  
Competition",  
RL Infect. Immun. 72:6151-6156 (2004).

DR EMBL; AY298765; AAQ75098.1; -; Genomic DNA.  
DR EMBL; AY299335; AAQ83784.1; -; Genomic DNA.  
DR EMBL; AY299336; AAQ83788.1; -; Genomic DNA.  
DR GO; GO:0016301; F:kinase activity; IEA.  
DR GO; GO:0004747; F:kinase activity; IEA.  
DR GO; GO:0016740; F:ribokinase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006014; P:D-ribose metabolism; IEA.  
DR InterPro; IPR011877; D:ribokin\_bact.  
DR InterPro; IPR011611; PfKB region.  
DR Pfam; PF00294; PfKB; 1.  
DR PRINTS; PR00990; RIBOKINASE.  
DR TIGRFAMs; TIGR02152; D\_ribokin\_bact; 1.  
KW Kinase.  
SQ SEQUENCE 306 AA; 33222 MW; 8A0FBF6BFE493DB CRC64;

Alignment Scores:  
Pred. No.: 1,34e-124 Length: 306  
Score: 1519.00 Matches: 294  
Percent Similarity: 98.69% Conservative: 8  
Best Local Similarity: 96.08% Mismatches: 4  
Query Match: 95.53% Indels: 0  
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q67FX7\_ECOLI (1-306)

QY 1 ATGGATATCGCGGTTATTGGCTCTAACATGGTGGACCTTACCTACCAACAGATG 60  
DB 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrTyrThrAsnGlnMet 20  
QY 61 CCCAAGAGGGAACTCTGAAGCGCGCGGTTTAAATCGGCTGGCGGGAAGGG 120  
DB 21 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyLysGly 40  
QY 121 CGGACACAGCGCTGGCGCGCTAAGCTCAATTCAAAGTATTGATGTTGACCAAGTG 180  
DB 41 AlaAsnGlnAlaValAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60  
QY 181 GCGCAGCATATTTTGGCCGACACACCATTCGTAATCTCGATCTGGGGATCAATACG 240  
DB 61 GlyAspAspIlePheAlaAspAsnThrIleArgAsnLeuGluSerTyrGlyIleAsnThr 80  
QY 241 AGGTATGTAGAAAAAGTACCGTGTACCGACGCGCGGTAGCGCGGATTTCTGCAACGCC 300  
DB 81 ThrTyrValGluLysValProCysThrSerSerGlyValAlaProIlePheValAsnAla 100  
QY 301 RACTCCAGCAAGCATCTGATCATCAAGGCGCTAACAGTTTCTCTCGCGCGGAAGAT 360  
DB 101 AsnSerSerAsnSerIleLeuIleIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120  
QY 361 ATCGATCGCGCGCGGAAGATTTAAAAATCCAGCTTATGTTCTGCAATCGGAAGTT 420  
DB 121 IleAspArgAlaAlaGluAspLeuLysCysLysLeuIleValLeuGlnLeuGluVal 140  
QY 421 CAGCTTGAACGGTTTATCACCAATAGAAATTGGCAAGAAACACAGGATTAAGTGTGA 480  
DB 141 GlnLeuGluThrValTyrHisAlaIleGluPheGlyLysLysAsnGlyIleGluValLeu 160  
QY 481 TTAACCTCGCCAGCATACGGGATAGATATGCTTATGCTGCTGTAATGCGGATTC 540  
DB 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyrAlaCysLysCysAspPhe 180  
QY 541 TTTGTACCTAATGAACAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGAC 600  
DB 181 PheIleProAsnGluThrGlnLeuGluIleLeuThrGlyMetSerValAspThrTyrAsp 200  
QY 601 CATATTTCGCGCAGCGCGCTTCGCTGGTAGATAAAGGCGCTGAACAATATTATGTCACC 660

Db 201 HisIleArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleValThr 220  
QY 661 ATGGCGCAGAAAGCGCGCTGTGGATGACGCGTGACAGGAAGTCCATGTTCCGCGCTTT 720  
Db 221 MetSerGluLysGlyAlaLeuTyrMetThrArgAspGlnGluValHisValProAlaPhe 240  
QY 721 AGACTGAACGCTGTGATACCGCGCGCGCGGCGGCGCTTTATCGGCTGTTTCGCGCAT 780  
Db 241 LysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 260  
QY 781 TACTACGTCCAGACGCGGCGATGTGAAAGCCGCCCATGAAAAAGCGCTCTCTTTGCCGCT 840  
Db 261 TyrTyrValGlnSerGlyAspValGluAlaAlaLeuLysLysAlaAlaLeuPheAlaAla 280  
QY 841 TTCAGCGTCAACCGGAAAGCAGCCCAATCTCTTATTCACGATTCAGCATTTAATGAG 900  
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyrProSerIleGluGlnPheAsnGlu 300  
QY 901 TATCTTTCGTTGAACGAA 918  
Db 301 PheLeuThrLeuAsnGlu 306

RESULT 6

Q8FKR4\_ECOL6 PRELIMINARY; PRT; 306 AA.

AC Q8FKR4; 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative ribokinase.

GN OrderedLocusNames=c0331;

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=217992;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=O6:H1 / CPT073 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,

RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;

RT "Extensive mosaic structure revealed by the complete genome sequence

RT of uropathogenic Escherichia coli.,"

RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

DR EMBL; AE016756; AAN78819.1; -; Genomic DNA.

DR HSP; P05054; 1RKD.

DR GO; GO:0016301; F:kinase activity; IEA.

DR GO; GO:0004747; F:ribokinase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR GO; GO:0006014; P:D-ribose metabolism; IEA.

DR InterPro; IPR011877; D:ribokin\_bact.

DR InterPro; IPR011611; PfKB region.

DR InterPro; IPR002139; Ribokinase.

DR Pfam; PF00294; PfKB; 1.

DR TIGRFAMs; TIGR02152; D\_ribokin\_bact; 1.

KW Complete proteome; kinase.

SQ SEQUENCE 306 AA; 33222 MW; 8A0FBF6BFE493DB CRC64;

Alignment Scores:

Pred. No.: 1,34e-124 Length: 306  
Score: 1519.00 Matches: 294  
Percent Similarity: 98.69% Conservative: 8  
Best Local Similarity: 96.08% Mismatches: 4  
Query Match: 95.53% Indels: 0  
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q8FKR4\_ECOL6 (1-306)

QY 1 ATGGATATCGCGGTTATTGGCTCTAACATGGTGGACCTTATCACCTACCAACAGATG 60

Db 1 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrThrAsnGlnMet 20  
Qy 61 CCCAAAGAGGGAAACTCTGAGACCGCGCGGTTTAAATCGCTCGCGGGAAGGG 120  
Db 21 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyVal 40  
Qy 121 CGCAACAGCGCGTGGCGCGCTAAAGCTCAATCAAAAGTATTGATGTTGACCAAGTG 180  
Db 41 AlaAsnGlnAlaValAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 60  
Qy 181 GCGCAGCATATTTTTCGCGACAACACCATTCGTATCTCGAATCTCGGGGATCAATACG 240  
Db 61 GlyAspAspIlePheAlaAsnThrIleArgAsnLeuGluSerTrpGlyLeuAsnThr 80  
Qy 241 ACGTATGTAGAAAAGTACCGTGTACAGCGCGCTAGCGCGATTTTCGTCAACGCC 300  
Db 81 ThrTyValGluLysValProCysThrSerGlyValAlaProIlePheValAsnAla 100  
Qy 301 AACTCCAGCAACAGCATTCGATCATCAAGCGCGCTACAAAGTTTCTCTCGCGGAAGAT 360  
Db 101 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 120  
Qy 361 ATCGATCCGCGCGCGAAGATTAAAAAATGCCAGCTTATTGTCTGCAACTGGAAGTT 420  
Db 121 IleAspArgAlaAlaGluAspLeuLysCysLysLeuIleValLeuGlnLeuGluVal 140  
Qy 421 CAGCTTGAAACCGTTTATCAGCAATAGAAATTTGGCAAGAAACACGGAATGAAGTGA 480  
Db 141 GlnLeuGluThrValTyHisAlaIleGluPheGlyLysLysAsnGlyIleGluValLeu 160  
Qy 481 TTAACCTCGCCAGCATTCAGGAATAGATATGCTTATGCTGTAAATGCGATTTC 540  
Db 161 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyAlaCysLysCysAspPhe 180  
Qy 541 TTTGTACCTAATGAAACCGAGCTGGAAATATTAAACCGGTATGCCAGTCGATACCTATGAC 600  
Db 181 PheIleProAsnGluThrGluLeuGluIleLeuThrGlyMetSerValAspThrTyAsp 200  
Qy 601 CATATTCGCGCAGCGCGCGTTCGTGTAGATAAAGCGGTGAACAATATTATGTCACC 660  
Db 201 HisIleArgLeuAlaAlaArgSerLeuValAspLysGlyLeuAsnAsnIleIleValThr 220  
Qy 661 ATGGCGAGAAAGCGCGCTGGATGACGGTGACCGAGGAGTCCATGTTCCGCGGTTT 720  
Db 221 MetSerGluLysGlyAlaLeuTrpMetThrArgAspGlnGluValHisValProAlaPhe 240  
Qy 721 AGAGTGAACGCTGTTGATACACGCGCGCGCGATGCTTTATCGGCTGTTTCGCGCAT 780  
Db 241 LysValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHis 260  
Qy 781 TACTACGTCCAGAGCGGGATGTGAAGCGCGCATGAAAAAGCGCTCTCTTTCGCGCT 840  
Db 261 TyrTyValGlnSerGlyAspValGluAlaLeuLysLysAlaAlaLeuPheAlaAla 280  
Qy 841 TTCAGCTCACCGGAAGGACCAATCTTATCCAGCATTCGCAATTTAATGAG 900  
Db 281 PheSerValThrGlyLysGlyThrGlnSerSerTyProSerIleGluGlnPheAsnGlu 300  
Qy 901 TATCTTTCTGTTGAACGAA 918  
Db 301 PheLeuThrLeuAsnGlu 306

RESULT 7

Q70712 ECOLI  
ID Q70712 ECOLI PRELIMINARY; PRT; 315 AA.  
AC Q70712  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=536;  
RX PubMed=15385503; DOI=10.1128/IAI.72.10.5993-6001.2004;  
RA Blum-Oehler G., Buchrieser C., Gottschalk G., Emody L., Hacker J.;  
"The Pathogenicity Island-Associated K15 Capsule Determinant Exhibits  
a Novel Genetic Structure and Correlates with Virulence in  
uropathogenic Escherichia coli Strain 536.";  
RL Infect. Immun. 72:5993-6001(2004).  
DR EMBL; AJ617685; CAB85173.1; -; Genomic DNA.  
DR GO; GO:0016301; F:kinase activity; IEA.  
DR GO; GO:0004747; F:ribokinase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006014; P:D-ribose metabolism; IEA.  
DR InterPro; IPR011877; D\_ribokin\_bact.  
DR InterPro; IPR011611; PfKB region.  
DR InterPro; IPR002139; Ribokinase.  
DR Pfam; PF00294; PfKB; 1.  
DR PRINTS; PR00990; RIBOKINASE.  
DR TIGRFAMs; TIGR02152; D\_ribokin\_bact; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 315 AA; 34327 MW; 8A9721EB93A15D59 CRC64;

Alignment Scores:

Pred. No.: 1.34e-124 Length: 315  
Score: 1519.00 Matches: 294  
Percent Similarity: 98.69% Conservative: 8  
Best Local Similarity: 96.08% Mismatches: 4  
Query Match: 95.53% Indels: 0  
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q70712\_ECOLI (1-315)

Qy 1 ATGATATTCGCGGTTATTGGCTTAACATGCTGGACCTTATCACCTACCAACCAAGATG 60  
Db 10 MetAspIleAlaValIleGlySerAsnMetValAspLeuIleThrThrAsnGlnMet 29  
Qy 61 CCCAAAGAGGGAAACTCTGAGACCGCGCGTTTAAATCGCTCGCGGGAAGGG 120  
Db 30 ProLysGluGlyGluThrLeuGluAlaProAlaPheLysIleGlyCysGlyGlyVal 49  
Qy 121 CGCAACAGCGCGTGGCGCGCTAAAGCTCAATCAAAAGTATTGATGTTGACCAAGTG 180  
Db 50 AlaAsnGlnAlaValAlaAlaLysLeuAsnSerLysValLeuMetLeuThrLysVal 69  
Qy 181 GCGCAGCATATTTTTCGCGACAACACCATTCGTATCTCGAATCTCGGGGATCAATACG 240  
Db 70 GlyAspAspIlePheAlaAsnThrIleArgAsnLeuGluSerTrpGlyLeuAsnThr 89  
Qy 241 ACGTATGTAGAAAAGTACCGTGTACCAAGCGCGCTAGCGCGATTTTCGTCAACGCC 300  
Db 90 ThrTyValGluLysValProCysThrSerGlyValAlaProIlePheValAsnAla 109  
Qy 301 AACTCCAGCAACAGCATTCGATCATCAAGCGCGCTACAAAGTTTCTCTCGCGGAAGAT 360  
Db 110 AsnSerSerAsnSerIleLeuIleLysGlyAlaAsnLysPheLeuSerProGluAsp 129  
Qy 361 ATCGATCCGCGCGCGAAGATTAAAAAATGCCAGCTTATTGTCTGCAACTGGAAGTT 420  
Db 130 IleAspArgAlaAlaGluAspLeuLysCysLysLeuIleValLeuGlnLeuGluVal 149  
Qy 421 CAGCTTGAAACCGTTTATCAGCAATAGAAATTTGGCAAGAAACACGGAATGAAGTGA 480  
Db 150 GlnLeuGluThrValTyHisAlaIleGluPheGlyLysLysAsnGlyIleGluValLeu 169  
Qy 481 TTAACCTCGCCAGCATTCAGGAATAGATATGCTTATGCTGTAAATGCGATTTC 540  
Db 170 LeuAsnProAlaProAlaLeuArgGluLeuAspMetSerTyAlaCysLysCysAspPhe 189



```

01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DN Sugar kinase.
GN OrderedLocusNames=V0528;
OS Vibrio vulnificus (strain VJ016).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=196600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=14656965; DOI=10.1101/gr.1295503;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RA "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen.";
RL Genome Res. 13:2577-2587(2003).
DR EMBL; BA000037; BAC93292.1; -; Genomic_DNA.
DR HSSP; P05054; 1GQT.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D:ribokin_bact.
DR InterPro; IPR011611; PfkB_region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfkB; 1.
DR PRINTS; PR00900; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
DR Complete proteome; Kinase.
KW SEQUENCE 309 AA; 33507 MW; EBB8ACF60CB6C44D CRC64;
SQ SEQUENCE 309 AA; 33507 MW; EBB8ACF60CB6C44D CRC64;

Alignment Scores:
Pred. No.: 4, 98e-91 Length: 309
Score: 1137.50 Matches: 218
Percent Similarity: 86.67% Conservative: 42
Best Local Similarity: 72.67% Mismatches: 39
Query Match: 71.54% Indels: 1
DB: 2 Gaps: 1

US-10-049-750-11 (1-921) x Q7MP36_VIBVY (1-309)
QY 7 ATCCGGTATTGGCTTAACATGGTGACCTTATCACCTACCAACGATGCCCAA 66
Db 4 IleAlaValIleGlySerAsnMetValAspLeuIleSerTyThrAspArgMetProlys 23
QY 67 GAAGGGAACTCGAAGCCGCGCTTTAAATCGCTCGCGGGAAGGGCGAC 126
Db 24 GluGlyGluThrLeuGluAlaProSerPheLysMetGlyCysGlyGlyLysGlyAlaAsn 43
QY 127 CAGGCGGCGGCGGCTTAAGCTCAATTCAAAAGTATTGATGTTGACCAAGTGGCGAC 186
Db 44 GlnAlaValAlaAlaAlaLysMetGlyAlaAspValValMetValSerLysValGlyAsp 63
QY 187 GATATTTTGGCACAACACCATTCGTATCTCGAATCTCGGGGATCAATACGACGTAT 246
Db 64 AspMetPheAlaAspAsnThrIleArgAsnPheGlnSerTyGlyIleAsnThrGlnTy 83
QY 247 GTAGAAAAGTACCGTGTACCGAGCGCGCTAGCGCGATTTTCGTCAACGCCCACTCC 306
Db 84 ValSerLysValProGlnThrSerSerGlyValAlaProIlePheValAsnSerThrSer 103
QY 307 AGCAACACATTCATCATCAAGCGCTCAACAGTTTCTCTCCCGGAAGATATCGAT 366
Db 104 GlnAsnSerIleLeuIleIleLysGlyAlaAsnGluPheLeuLysProAspAspIleAsp 123
QY 367 CGCGCGGCGGAGATTAATAAATGCGCATTTATTTGTTGCAATCGAAGTTCAGCTT 426
Db 124 LysAlaGluSerThrLeuValGluCysSerLeuIleValLeuGlnLeuGluValProLeu 143
QY 427 GAAACGGTTTATCAGCAATAGAAATTTGGCAAGAACACGGGATTCAGTGTATTAAAC 486
Db 144 GluThrValTyAlaAlaIleGluPheGlyAsnLysHisSerIleProValLeuLeuAsn 163
QY 487 CTGCGCGCCAGCATTACGGGAATTAGATATGCTTATGCTGTAAATGCGATTTCTTTGTA 546

Db 164 ProAlaProAlaValProGluLeuAspIleGluTyAlaCysArgCysAspPheVal 183
QY 547 CCTAATGAAACCGAGCTGGAAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
Db 184 ProAsnGluThrGluLeuGluIleLeuValAsnLysProValGluThrValGluGlnIle 203
QY 607 CGCGCAGCGCACGTTCCGCTGGTAGATAAAGGCTGAACCAATATTATTGTCCACCATGGC 666
Db 204 LysGluAlaAlaThrIleLeuLeuAsnLysGlyLeuAsnAsnIleIleValThrMetGly 223
QY 667 GAGAAAGCGCGCTGTGGATGACCGCGTGAC---CAGGAAGTCCATGTTCCGGCGTTTAGA 723
Db 224 SerLysGlyAlaLeuTrpLeuSerLysAspGlyLysAspValPheIleGluProThrLys 243
QY 724 GTGAACCTGTTGATACACGCGCGCGGCGGATGCTTTATCGCTGTTTCGCGCATTTAC 783
Db 244 ValAsnAlaValAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheSerHisTy 263
QY 784 TACGTCAGAGCGGGGATGTGAAGCGCCCATGAAAAAGCCGTCCTCTTTGCCGCTTTC 843
Db 264 PheMetGlnThrGlyAspValGlnLysSerLeuGluLysAlaSerLeuPheAlaAlaPhe 283
QY 844 AGCTCACCGGGAAGCAGCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGAGTAT 903
Db 284 SerValThrGluLysGlyThrGlnPheSerTyThrGlnPheSerIleGluGlnPheGluGluPhe 303

RESULT 10
Q6ALS2_DESPS
ID Q6ALS2 DESPS PRELIMINARY; PRT; 309 AA.
AC Q6ALS2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Related to ribokinase.
GN OrderedLocusNames=DPI974;
OS Desulfotalea psychrophila.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobacteriales;
OC Desulfobulbaceae; Desulfotalea.
OX NCBI_TaxID=84980;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=LSV54 / DSM 12343;
RX PubMed=15305914; DOI=10.1111/j.1462-2920.2004.00665.x;
RA Rabus R., Ruepp A., Frickey T., Rattei T., Fartmann B., Stark M.,
RA Bauer M., Zibat A., Lombardot T., Becker I., Amann J., Gellner K.,
RA Teeling H., Leuschner W.D., Gloeckner F.-O., Lupas A.N., Amann R.,
RA Klenk H.-P.;
RT "The genome of Desulfotalea psychrophila, a sulfate-reducing bacterium
RT from permanently cold Arctic sediments.";
RL Environ. Microbiol. 6:887-902(2004).
DR EMBL; CR522870; CAG36703.1; -; Genomic_DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D_ribokin_bact.
DR InterPro; IPR011611; PfkB_region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfkB; 1.
DR PRINTS; PR00900; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
KW Complete proteome; Kinase.
SQ SEQUENCE 309 AA; 32997 MW; 4A16AD3CC4C86267 CRC64;

Alignment Scores:
Pred. No.: 1, 36e-72 Length: 309
Score: 928.00 Matches: 183
Percent Similarity: 74.92% Conservative: 41
Best Local Similarity: 61.20% Mismatches: 75
Query Match: 58.36% Indels: 0
DB: 2 Gaps: 0
```

US-10-049-750-11 (1-921) x Q6ALS2\_DESPS (1-309)

```

QY 7 ATCCGGGTTATTGGCTTAACATGTGGACCTTATCACCACCAACAGATGCCAAA 66
Db 7 lIeAlaValIleGlySerAsnMetValAspLeuValThrTyrThrAspArgMetProVal 26

QY 67 GAAGGGGAAACTCTGGAAGCGCGGCTTTAAATATCGGCTGGCGGGAAGGGCGAAC 126
Db 27 AlaGlyGluThrLeuGluAlaProAspPheAspLeuGlyPheGlyGlyLysGlyAlaAsn 46

QY 127 CAGCGCGTGGCGCGCTTAAGCTCAATCAAAAGTATTGATCTGACCAAGTGGCGCAC 186
Db 47 GlnAlaIleAlaAlaLysLeuGlyGlyGluValLeuMetLeuSerTyrValGlyAsp 66

QY 187 GATATTTTTCGCGACACACCATTCGTATCTCGAATCTCGATCTGCGGATCAATACGACGTAT 246
Db 67 AspValPheGlyProAsnThrArgAlaAsnPheValAlaAsnGlyIleAspAlaArgTyr 86

QY 247 GTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGATTTTCGTCAACGCCCACTCC 306
Db 87 ValGluThrAlaAlaGlyLeuSerSerGlyValAlaProIlePheValAspAlaLysGly 106

QY 307 AGCAACAGCATTCGATCATCAAGCGCTTAACAAGTTCTCTCGCGGAGATATCGAT 366
Db 107 GlnAsnSerIleLeuIleValLysGlyAlaAsnLysHisLeuSerProValAspValAsp 126

QY 367 CGCGCGCGGAGAGATTAAATAATCCAGCTTATTGTTCTCAACTGGAAGTTCAGCTT 426
Db 127 ArgAlaIleAspIleArgAlaCysAspLeuIleValMetGlnLeuGlyIleSerLeu 146

QY 427 GAAACGGTTTATCAGCAATAGATATTGGCAAGAAACACGCGGATTCAGTGTATTAAAC 486
Db 147 GluThrValTyrTyrValIleGluLeuGlyLeuArgHisIleProValLeuLeuAsn 166

QY 487 CTGCGCCAGCATTCGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
Db 167 ProAlaProAlaValAlaGlyLeuAspMetGluLysIleCysGlnLeuAspMetLeuVal 186

QY 547 CCTAATGAACCGAGTGGAAATATTAAACCGTATGCCAGTGGATACCTATCACCATT 606
Db 187 ProAsnGluThrGluLeuGluIleLeuThrGlyMetProValGlnThrLeuGluGlnIle 206

QY 607 CGCGCAGCGCGCGTTCGCTGTAGATAAAGGCGTGAACAATATTATTGTCACCATGGGC 666
Db 207 GlnAlaAlaAlaLysPheLeuIleAspAlaGlyIleLysLysValIleValThrMetGly 226

QY 667 GAGAAAGCGCGCTGTGGTAGCGGTGACCGAGAGTCCATGTTCCGCGGTTTAGGTG 726
Db 227 SerLysGlyAlaLeuLeuValThrGluLysGluMetThrSerValProCysProGlnVal 246

QY 727 AACGCTGTGTATACCGCGCGCGGTGATGCTTTATCGGCTGTTTCGCGCATCTACTAC 786
Db 247 LysAlaLysAspThrSerGlyAlaGlyAspAlaPheIleGlyCysPheAlaLysHisTyr 266

QY 787 GTCCAGAGCGGAGTGTGAAGCGCGCATGAAAGCGCTCTCTTTCCGCTTTTCAGC 846
Db 267 ValGluGlyGlyGluLeuIleProAlaMetGluGluAlaValIleTyrAlaSerLeuSer 286

QY 847 GTCACGGGAAAGGCAACCAATCTCTTATCCAAGCATTTGAGCAATTTAATGAGTAT 903
Db 287 ThrThrArgProGlyThrGlnLysSerTyrAlaAspIleAspGlnPheGluAlaTyr 305

```

# RESULT 11

Q4LSV5\_9BURK PRELIMINARY; PRT; 320 AA.

AC Q4LSV5;

DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Carbohydrate kinase, PfKB.

GN ORFNames=Bcen2424DRAFT\_3825;

OS Burkholderia cenocepacia HI2424.

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia; Burkholderia cenocepacia complex.  
 OX NCBI\_TaxID=331272;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=HI2424;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,  
 Hammon N., Istrati S., Pittluck S., Richardson P.;  
 RT "Sequencing of the draft genome assembly of Burkholderia cenocepacia  
 HI2424.";  
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=HI2424;  
 RG US DOE Joint Genome Institute (JGI-ORNL);  
 RA Larimer F., Land M.;  
 RT "Annotation of the draft genome assembly of Burkholderia cenocepacia  
 HI2424.";  
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 DR EMBL; AHU01000019; EMBL19121.1; -; Genomic\_DNA.  
 KW Kinase.  
 SQ SEQUENCE 320 AA; 33910 MW; 9A1441345CD3FA07 CRC64;

## Alignment Scores:

Align. No.: 3,076-72 Length: 320  
 Score: 924.00 Matches: 183  
 Percent Similarity: 75.33% Conservatives: 43  
 Best Local Similarity: 61.00% Mismatches: 74  
 Query Match: 58.11% Indels: 0  
 DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q4LSV5\_9BURK (1-320)

```

QY 7 ATCCGGTATTGGCTTAACATGTGGACCTTATCACCACCAACAGATGCCAAA 66
Db 16 lIeAlaValIleGlySerAsnMetValAspLeuValThrTyrValAlaArgMetProAla 35

QY 67 GAAGGGGAAACTCTGGAAGCGCGCTTTAAATCGGCTGGCGGGAAGGGCGAAC 126
Db 36 ArgGlyGluThrLeuGluAlaProAsnPheGluLeuGlyCysGlyLysGlyAlaAsn 55

QY 127 CAGCGCGTGGCGCGCTTAAGCTCAATCAAAAGTATTGATCTTACCAAGTGGCGCAC 186
Db 56 GlnAlaValAlaAlaAlaArgLeuGlyAlaArgValValMetValThrLysValGlyAsp 75

QY 187 GATATTTTTCGCGACACACCATTCGTATCTCGAATCTCTGGAATCTCTGGGGATCAATACGACGTAT 246
Db 76 AspValPheAlaAspAsnThrIleArgAsnPheGluArgGluGlyIleAspThrThrHis 95

QY 247 GTAGAAAAGTACCGTGTACACGACGCGGTAGCGCGATTTTCGTCAACGCCCACTCC 306
Db 96 ValArgLysValAlaGlyValProSerGlyValAlaProIlePheValGluProAspSer 115

QY 307 AGCAACAGCATTCGATCATCAAGCGCGCTTAACAGTTCCTCTCGCGGAGAGATATCGAT 366
Db 116 SerAsnSerIleLeuIleValLysGlyAlaAsnArgHisLeuGlnProAlaAspIleAsp 135

QY 367 CGCGCGCGGAGAGATTTAAATAATGCCAGCTTATTGTTCTGCAACTGGAAGTTCAGCTT 426
Db 136 AlaAlaAlaProMetLeuAlaGluCysAlaLeuIleValLeuGlnLeuGluLeu 155

QY 427 GAAACGGTTTATCAGCAATAGATATTGGCAAGAAACACGCGGATTCAGTGTATTAAAC 486
Db 156 AspThrValTyrHisAlaIleGluPheGlyAlaArgHisGlyIleProValLeuLeuAsn 175

QY 487 CTGCGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGCTGTAATGCGATTTCTTTGTA 546
Db 176 ProAlaProAlaValAlaAspLeuAspPheGluArgIleArgSerValGluPheVal 195

```



RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=K96243;  
RX PubMed=15377794; DOI=10.1073/pnas.0403201101;  
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdeno-Tarriaga A.-M.,  
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,  
RA Bentley S.D., Sebahia M., Thomson N.F., Bason N., Beacham I.R.,  
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,  
RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshazer D.,  
RA Feltwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagels K.,  
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,  
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,  
RA Songsivilai S., Stevens K., Tumapa S., Vesaratchavest M.,  
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;  
RA "Genomic plasticity of the causative agent of melioidosis,  
RT Burkholderia pseudomallei";  
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245 (2004).  
DR EMBL; BX571966; CAH38316.1; -; Genomic DNA.  
DR GO; GO:0016301; F:kinase activity; IEA.  
DR GO; GO:0004747; F:ribokinase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006014; P:D-ribose metabolism; IEA.  
DR InterPro; IPR011877; D:ribokin\_bact.  
DR InterPro; IPR011611; PfkB\_region.  
DR InterPro; IPR002139; Ribokinase.  
DR Pfam; PF00294; PfkB; 1.  
DR PRINTS; PR00990; RIBOKINASE.  
DR TIGRFAMs; TIGR02152; D\_ribokin\_bact; 1.  
DR Complete proteome; Kinase.  
KW TIGRPFAMs; TIGR02152; D\_ribokin\_bact; 1.  
SQ SEQUENCE 310 AA; 32606 MW; A58CAE50FC6754C CRC64;

Alignment Scores:  
Pred. No.: 28-69 Length: 310  
Score: 892.00 Matches: 170  
Percent Similarity: 76.00% Conservative: 58  
Best Local Similarity: 55.67% Mismatches: 72  
Query Match: 56.10% Indels: 0  
Gaps: 0

US-10-049-750-11 (1-921) x Q63M06\_BURPS (1-310)

QY 7 ATCGCGGTTATGGCTTAACATGGTGGACCTTATCACCACCAACAGATGCCAAA 66  
Db 4 ileAlaValIleGlySerAsnMetValAspLeuValThrTyValThrArgMetProAla 23  
QY 67 GAAGGGGAAACTCTGGAACCGCGCGCTTTAAATCGGTCGGCGGAAAGGGCGAAC 126  
Db 24 AspGlyGluThrLeuGluAlaProAsnPheGluLeuGlyCysGlyGlyAlaAsn 43  
QY 127 CAGGCGGTCGGCGCGCTAAGCTCAATTCAAAAGTATTGATGTGACCAAGTGGCGAC 186  
Db 44 GlnAlaValAlaAlaSerLysLeuGlyAlaArgValAlaMetIleSerLysLeuGlyAsp 63  
QY 187 GATATTTTTCGCGCAACACCATTCGTATCTCGAATCCTGGGGATCAATACGATAT 246  
Db 64 AspLeuPheAlaGluAsnThrLeuArgAsnPheGluArgPheGlyValAspThrGluHis 83  
QY 247 GTAGAAAAGTACCGGTACCGAGCGCGGTAGCGCCGATTTTGTCAACCCCACTCC 306  
Db 84 ValArgArgValSerGlyValSerGlyValAlaProIlePheValSerProAspSer 103  
QY 307 AGCAACAGCATTCGTGATCATCAAGCGCTACAGTTCCTCGCGCGGAAGATATCGAT 366  
Db 104 ArgAsnArgIleLeuIleValLysGlyAlaAsnArgHisLeuArgProAlaAspIleAsp 123  
QY 367 CGCGCGCGGAAGATTAAATAAATCCAGCTTATTTCTGCAATCGAAGTTCAGCTT 426  
Db 124 AlaAlaAlaAlaLysIleGluAlaSerArgLeuValValLeuGlnLeuGluIleAspIle 143  
QY 427 GAAACGGTTTATCAGCAATAGATTTGGCAAGAAACACAGGATGAAGTGTATTAAAC 486  
Db 144 AspThrValTyThrAlaIleAspPheAlaAlaAlaArgGlyIleProValLeuLeuAsn 163

QY 487 CCTGGCGCAGCATTTACGGGAATTAGATATGCTTATGCTGCTAAATGCGATTCTTCTGA 546  
Db 164 ProAlaProGlyValProAspLeuAspPheAlaArgLeuAlaLysLeuGluPheLeuVal 183  
QY 547 CCTAATGAACCGAGCTCGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606  
Db 184 ProAsnGluThrGluLeuAlaLeuValSerGlyMetProThrAspThrProAspAlaVal 203  
QY 607 CGCGCAGCGGACGTTCTGCTGGTGTAGATAAAGGCTGAACAATATTATTGTACCATGGCG 666  
Db 204 GluArgAlaAlaGlySerLeuValGluArgGlyValLysHisValIleValThrLeuGly 223  
QY 667 GAGAAAGCGCGCTGTGATCAGCGTGACCGAGTCCATGTTTCCGCGGTTTAGAGTG 726  
Db 224 GluLysGlySerLeuLeuValSerArgAlaGlyAlaValArgValProProValAlaVal 243  
QY 727 AACGCTGTTGTATACAGCGCGCGCGGATCCCTTTATCGGCTGTTTCCGCGATTACTAC 786  
Db 244 AspAlaArgAspThrThrGlyAlaGlyAspAlaTyIleGlyCysPheAlaArgHisTy 263  
QY 787 GTCCAGAGCGGGGATGTGGAAGCGCGCATGAAAGCGGCTCTCTTTGCGCGCTTTCAGC 846  
Db 264 ValAlaThrAlaAspIleProAlaAlaMetArgLeuAlaSerAlaTyAlaAlaHisSer 283  
QY 847 GTCACCGGGAAGCGACCCCATCTCTTATCCAGCATTTGAGCATTTAAATGAGTATCTT 906  
Db 284 ValThrGlyLeuGlyThrGlnLysSerTyAlaAlaAlaThrPheLeuArgPheLeu 303

RESULT 14  
Q8U6G0\_AGR75 PRELIMINARY; PRT; 312 AA.  
AC Q8U6G0; Q7CW10;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Ribokinase (AGR\_L78p).  
GN Name=rbks; OrderedLocNames=AGR\_L78, Atu4847;  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Agrobacterium.  
OX NCBI\_TaxID=176299;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D. Sr.,  
RA Chapman P., Glendinning J., Deatherage G., Gillet W., Grant C.,  
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,  
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Perry M.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58";  
RT Science 294:2317-2323 (2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,  
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Houmle K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58";  
RT Science 294:2323-2328 (2001).  
RL EMBL; AE009413; AAL45641.1; -; Genomic DNA.  
DR EMBL; AE008201; AK88602.1; -; Genomic DNA.  
DR PIR; AC3153; AC3153.  
DR PIR; H98134; H98134.



```
DR HSSP; P05054; IRKD.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D:ribokin_bact.
DR InterPro; IPR011611; PfKB region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfKB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
KW Complete proteome; Kinase.
SQ SEQUENCE 312 AA; 33094 MW; D2A070A3549E35F1 CRC64;

Alignment Scores:
Pred. No.: 3,41e-68 Length: 312
Score: 878.00 Matches: 168
Percent Similarity: 76.25% Conservativeness: 60
Best Local Similarity: 56.19% Mismatches: 71
Query Match: 55.22% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-11 (1-921) x Q8U6G0_AGR75 (1-312)
QY 7 ATCCGGGTATTGGCTTAACATGCTGGACCTTATCACCCTACACCAACAGATGCCCAA 66
DB 5 IleGlyValValGlySerAsnMetValAspLeuIleThrTyValAspArgMetProgly 24
QY 67 GAAGGGAACTCGGAAGCCGCGCTTTAAATCGCTCGCGGGAAGAGGGCGAAC 126
DB 25 ProGlyGluThrLeuGluAlaProThrPheGluMetGlyCysGlyGlyLeuAlaAsn 44
QY 127 CAGCGCGGGCGCGCTAAGCTCAATCAAAAGTATTGATTGACCAAGTGGCGAC 186
DB 45 GlnAlaValAlaAlaAlaArgLeuGlyAlaGluValMetMetValThrArgValGlyAsp 64
QY 187 GATATTTTGGCGCAACACCATTCGTAATCTCGAATCTCGGGGATCAATACACGTTAT 246
DB 65 AspValPheAlaAspAsnThrIleArgAsnLeuAlaSerPheGlyValAspThrArgHis 84
QY 247 GTAGAAAAGTACCGTGTACGAGCGCGGTAGCGCGATTTCTCGTCAACGCCCACTCC 306
DB 85 ValValIysValSerGlyLysSerSerGlyValAlaProIlePheValGluGlnSerGly 104
QY 307 AGCAACACATTCGATCATCAAGAGCGCTAACAGTTTCTCGCGGAGAGATATCAT 366
DB 105 GluAsnSerIleLeuIleValIysGlyAlaAsnAlaAspLeuLeuProValGluValAsp 124
QY 367 CGCGCGCGGAGAGATTTAAAAAATGCCAGCTTATTGTTCTGCAACTGGAAAGTTCAGCTT 426
DB 125 LysAlaAlaAlaAspLeuLysGluCysGlyLeuIleLeuMetGlnMetGluValProVal 144
QY 427 GAAACGGTTTATCGCAATAGAAATTTGGCAAGAAACACGGGATTGAAGTGTATTAAAC 486
DB 145 GluThrValTyHisThrIleGluPheAlaAlaGlnAsnGlyIleGluThrIleLeuAsn 164
QY 487 CTGCGCCAGCATTACGGGAATTAGATGCTTATGCTTATGCTGCTTAATGCGATTTCTTGTGA 546
DB 165 ProAlaProAlaAlaAlaAsnLeuAspProGluArgIleArgGlnValThrPheLeuVal 184
QY 547 CTTAATGAACCGAGCTGGAATATTAAACCGGTATGCCAGTGGATACCTATGACCATATT 606
DB 185 ProAsnGluSerGluLeuAlaLeuLeuSerGlyLeuProThrAspThrAspGluAspIle 204
QY 607 CGCGCAGCGGCGCTGCTGCTAGATAAAGGCTGAACAATATTATTGTTCACCATGGGC 666
DB 205 ValArgAlaAlaArgSerLeuIleAlaArgGlyIleArgThrValIleValThrLeugly 224
QY 667 GAGAAAGCGCGCTGTGATCGCTGACCGGTGACAGGAAGTCCATGTTCCGGCGTTTAGAGTG 726
DB 225 AlaArgGlyAlaArgMetIleThrSerAspGluIleValAsnIleGluProValIysVal 244
QY 727 AACGCTGTTGATACAGCGCGCGCGATGCTTTATCGGCTGTTTTCGCGCATTTACTTAC 786
DB 727 AACGCTGTTGATACAGCGCGCGCGATGCTTTATCGGCTGTTTTCGCGCATTTACTTAC 786

Db 245 ThrProArgAspThrThrGlyAlaGlyAspAlaPheIleGlySerPheAlaArgPheTy 264
QY 787 GTCCAGAGCGGGATGTGAAGCCGCATGAAAAAGCCGCTCTCTTTGCCGCTTTTCAGC 846
DB 265 AlaGluThrArgGluValValSerSerLeuLysLysAlaSerLeuTyAlaAlaHisSer 284
QY 847 GTCACCGGGAAGACCCCAATCTCTTATCCAGCATTTGAGCAATTTAATGAGTAT 903
DB 285 IleThrArgProGlyThrGlnLysAlaTyAlaSerIleAspGluPheGluAlaPhe 303

RESULT 15
Q6A8F4_PROAC
ID Q6A8F4_PROAC PRELIMINARY; PRT; 304 AA.
AC Q6A8F4;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Sugar kinase, ribokinase family (EC 2.7.1.15).
GN OrderedLocusNames=PPA1211;
OS Propionibacterium acnes.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Propionibacterineae; Propionibacteriaceae; Propionibacterium.
OX NCBI_TaxID=11477;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=KPA171202 / DSM 16379;
RX PubMed=15286373; DOI=10.1126/science.1100330;
RA Brueggemann H., Henne A., Hoster F., Liesegang H., Wiewer A.,
RA Strittmatter A., Hujer S., Duerre P., Gottschalk G.;
RT "The complete genome sequence of Propionibacterium acnes, a commensal
of human skin.";
RL Science 305:671-673 (2004).
DR EMBL; AE017283; AAT82961.1; -, Genomic DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR011877; D_ribokin_bact.
DR InterPro; IPR011611; PfKB region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfKB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR TIGRFAMs; TIGR02152; D_ribokin_bact; 1.
KW Complete proteome; Kinase; Transferase.
SQ SEQUENCE 304 AA; 32566 MW; 3C03C7FA5FAE2080 CRC64;

Alignment Scores:
Pred. No.: 4,19e-64 Length: 304
Score: 831.50 Matches: 168
Percent Similarity: 73.97% Conservativeness: 48
Best Local Similarity: 57.53% Mismatches: 75
Query Match: 52.30% Indels: 1
DB: 2 Gaps: 1

US-10-049-750-11 (1-921) x Q6A8F4_PROAC (1-304)
QY 1 ATGATATCGCGGTATTGGCTTAACATGCTGGACCTTATCACCCTACCAACCAAGATG 60
DB 1 MetAspIleAlaValValGlySerAsnMetValAspLeuIleSerTyIleHisArgMet 20
QY 61 CCCAAAGAGGGGAACTCTGGAAGCGCGCGTGTAAATCGGCTCGCGGGAAGGG 120
DB 21 ProSerAspGlyGluThrValGluAlaProGluPheArgMetGlyCysGlyLysGly 40
QY 121 GCGAACACAGCGCGTGGCGCGCTAAGCTCAATCAAAAGTATTGATTTGACCAAGTG 180
DB 41 AlaAsnGlnAlaValAlaAlaSerArgLeuGlyAlaGluValValMetValThrArgVal 60
QY 181 GCGCAGCATTTTTCGCGCAACCAACATTCGTAATCTCGAATCTCGGCGGATCAATACG 240
DB 61 GlyAsnAspValPheAlaAspThrThrLeuAspAsnPheArgLysAsnGlyLysThr 80
QY 241 ACGTATGTAGAAAAAGTACCGTGTACCAGCAGCGCGTAGCGCGGATTTTCGTCAACGCC 300
```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:05:55 ; Search time 90.6485 Seconds  
(without alignments)  
4682.260 Million cell updates/sec

Title: US-10-049-750-13

Perfect score: 861

Sequence: 1 gttactactaaatgccaaaaa.....acgaagggtgcctatactaa 483

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cpn2.1/USPTO.spool.p/US10049750/runat.23122005.113551.6285/app.query.fasta.1.1742  
-DB=A\_Geneseq -QFM=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LCORCL=0 -LOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10049750 @CGN 1 1 476 @runat.23122005.113551.6285 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq.21.\*  
1: Genesep1980s.\*  
2: Genesep1990s.\*  
3: Genesep2000s.\*  
4: Genesep2001s.\*  
5: Genesep2002s.\*  
6: Genesep2003as.\*  
7: Genesep2003bs.\*  
8: Genesep2004s.\*  
9: Genesep2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	848	98.5	157	AAB67589	Aab67589 Amino aci
2	842	97.8	157	ADR97103	Adr97103 N-deoxyri
3	728	84.6	158	ABP98175	Abp98175 Amino aci
4	719	83.5	159	ABP98180	Abp98180 Amino aci
5	637	74.0	158	AAG66170	Aag66170 L. helvet
6	629	73.1	133	ABP98179	Abp98179 Amino aci
7	405	47.0	84	ABP98178	Abp98178 Amino aci
8	209	24.3	168	ABP98177	Abp98177 Amino aci
9	166.5	19.3	167	ABP98176	Abp98176 Amino aci

10	142	16.5	216	3	AAB11729	Aab11729 Cryptospo
11	142	16.5	216	5	ABJ04047	Abj04047 C parvum
12	142	16.5	1837	3	AAB11726	Aab11726 Cryptospo
13	142	16.5	1837	5	ABJ04044	Abj04044 C parvum
14	138.5	16.1	249	3	AAB11746	Aab11746 C parvum
15	138.5	16.1	249	5	ABJ04059	Abj04059 C parvum
16	138.5	16.1	1721	2	AAB48299	Aab48299 Cryptospo
17	138.5	16.1	1721	2	AAB48299	Aab48299 Cryptospo
18	136.5	15.9	159	5	ABJ04045	Abj04045 C parvum
19	136.5	15.9	159	5	ABJ04045	Abj04045 C parvum
20	134.5	15.6	175	5	ABJ04056	Abj04056 C parvum
21	134.5	15.6	175	5	ABJ04056	Abj04056 C parvum
22	132	15.3	216	3	AAB11734	Aab11734 Cryptospo
23	131	15.2	343	5	ABP69313	Abp69313 Human pol
24	128	14.9	2624	8	ADP97567	Adp97567 Pancreat
25	126.5	14.7	288	2	AAY29082	Aay29082 T. gondii
26	126.5	14.7	288	2	AAY29081	Aay29081 T. gondii
27	126.5	14.7	288	4	AU25553	Au25553 T. gondii
28	126.5	14.7	288	4	AU25552	Au25552 T. gondii
29	126.5	14.7	288	7	ADG17391	Adg17391 T. gondii
30	126.5	14.7	288	7	ADG17394	Adg17394 T. gondii
31	126.5	14.7	341	7	ABO73761	Ab073761 Pseudomon
32	126	14.6	1795	4	ABB69806	Abb69806 Drosophil
33	125.5	14.6	605	8	ADS43586	Ads43586 Bacterial
34	124	14.4	292	8	ADX96526	Adx96526 Plant ful
35	124	14.4	580	8	ADN72473	Adn72473 Thale cre
36	120.5	14.0	618	7	ABO79612	Ab079612 Pseudomon
37	119.5	13.9	618	7	ABO74950	Ab074950 Pseudomon
38	119	13.8	281	8	ABO84615	Ab084615 Human can
39	119	13.8	324	8	ADX90939	Adx90939 Plant ful
40	119	13.8	347	8	ADX93820	Adx93820 Plant ful
41	119	13.8	361	8	ADX96938	Adx96938 Plant ful
42	119	13.8	362	8	ADY12362	Ady12362 Plant ful
43	117.5	13.6	333	8	ADY05763	Ady05763 Plant ful
44	117.5	13.6	383	8	ADX90993	Adx90993 Plant ful
45	117.5	13.6	398	8	ADX77255	Adx77255 Plant ful

#### ALIGNMENTS

RESULT 1

AAB67589

ID AAB67589 standard; protein; 157 AA.

XX

AC AAB67589;

XX

DT 29-MAY-2001 (first entry)

XX

DE Amino acid sequence of a nucleoside 2-deoxyribosyltransferase.

XX

KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;  
KW purine nucleoside phosphorylase; phosphopentose mutase;  
KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;  
KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.

XX

OS Lactobacillus leichmannii.

XX

PN WO200114566-A2.

XX

PD 01-MAR-2001.

XX

PF 18-AUG-2000; 2000WO-EP008088.

XX

PR 20-AUG-1999; 99EP-00116425.

XX

PA (HOFF ) ROCHE DIAGNOSTICS GMBH.

XX

PA (INSP ) INST PASTEUR.

XX

PA (PHAR-) PHARMA-WALDHOFF GMBH & CO KG.

XX

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik B;

XX

PI Marliere P, Pochet S;

DR

WPI; 2001-235026/24.

DR N-PSDB; AAF55445.  
 XX  
 PT In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting  
 PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside  
 PT and an inorganic phosphate.  
 XX  
 XX  
 PS Disclosure; Page 62-63; 73pp; English.  
 XX  
 CC The present sequence represents a nucleoside 2-deoxyribosyltransferase  
 CC enzyme. This enzyme is involved in the biosynthesis of  
 CC deoxyribonucleosides, and is used in the method of the invention. The  
 CC specification describes a method for the in vitro enzymatic synthesis of  
 CC deoxyribonucleosides. The method comprises reacting deoxyribose 1-  
 CC phosphate and a nucleobase to form a deoxyribonucleoside and an inorganic  
 CC phosphate. Enzymes which may be used in the method of the invention  
 CC include thymidine phosphorylase, purine nucleoside phosphorylase,  
 CC phosphopentose mutase, phosphopentose aldolase, fructose 1,6-diphosphate  
 CC aldolase, deoxyribokinase, and nucleoside 2-deoxyribosyltransferase  
 XX  
 SQ Sequence 157 AA;

Alignment Scores:  
 Pred. No.: 3,11e-80 Length: 157  
 Score: 848.00 Matches: 157  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 98.49% Indels: 0  
 DB: 4 Gaps: 0

US-10-049-750-13 (1-483) x AAB67589 (1-157)

QY 10 ATGCCAAAAGACGATCTACTTCGGTGGCGGTGGTCTCACTGACGCCCAAAACAAAGCC 69  
 Db 1 MetProLysLysThrIleTyrPheGlyAlaGlyTrpPheThrAspArgGlnAsnLysAla 20  
 QY 70 TACAAGGAGCCATGAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAACACGCTAC 129  
 Db 21 TyrLysGluAlaMetGluAlaLeuLysGluAsnProThrIleAspLeuGluAsnSerTyr 40  
 QY 130 GTTCCCTCGACCAACGATACAGGGTATCCGGTTTGATGAACACCCGGAATACCTGCAT 189  
 Db 41 ValProLeuAspAsnGlnTyrLysGlyIleArgValAspGluHisProGluTyrLeuHis 60  
 QY 190 GACAAGGTTGGCTAGCGCCACCTACAAACGACTTGAACGGGATCAAGACCAACGAC 249  
 Db 61 AspLysValTrpAlaThrAlaThrTyrAsnAsnAspLeuAsnGlyIleLysThrAsnAsp 80  
 QY 250 ATCATGCTGGTGTCTACATCCCTGACGAAGACGTCGGCTGGGCATGGAACTGGGT 309  
 Db 81 IleMetLeuGlyValTyrIleProAspGluGluAspValGlyLeuGlyMetGluLeuGly 100  
 QY 310 TAGCCCTTGAGCAAGCAAGTACGTCCTTTTGGTCATCCCGGACGAAGACTACGCAAG 369  
 Db 101 TyrAlaLeuSerGlnGlyLysTyrValLeuLeuValIleProAspGluAspTyrGlyLys 120  
 QY 370 CCGATCAACCTCATAGCTGGCGCTCAGCGCAACAGTGATCAAGATGACCGACGCTGAAG 429  
 Db 121 ProIleAsnLeuMetSerTrpGlyValSerAspAsnValIleLysMetSerGlnLeuLys 140  
 QY 430 GACTTCAACTTCAACAGCCGCTTCGACTTCTACGAAGTCCCGCTATAC 480  
 Db 141 AspPheAsnPheAsnLysProArgPheAspPheTyrGluGlyAlaValTyr 157

RESULT 2

ADR97103  
 ID ADR97103 standard; protein; 157 AA.

AC ADR97103;

DT 16-DEC-2004 (first entry)

DE N-deoxyribosyltransferase (DTP) with G9S mutant, seq id 1.

XX

KW Cytostatic; antimicrobial; N-dideoxyribosyltransferase; DDTp;  
 KW N-deoxyribosyltransferase; DTP; antitumour; infection; ddc; ddi; muten.  
 XX  
 OS Lactobacillus leichmannii.  
 OS Synthetic.  
 XX

Key Location/Qualifiers  
 FT Misc-difference 9 /note= "Wild-type Gly replaced with Ser"  
 FT

XX FR2852968-Al.

XX 01-OCT-2004.

XX 28-MAR-2003; 2003FR-00003910.

XX 28-MAR-2003; 2003FR-00003910.

XX (INSP ) INST PASTEUR.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Kaminski PA;

XX WPI; 2004-693069/68.

XX N-PSDB; ADR97104.

XX Method for evolution of proteins to modify characteristics, useful  
 PT particularly for evolving an N-deoxyribosyltransferase to produce an N-  
 PT dideoxyribosyltransferase for preparation of antitumor nucleosides and  
 PT nucleotides.

XX Claim 15; SEQ ID NO 2; 51pp; French.

XX The invention relates to a method for evolution of a protein (X) so as to  
 CC modify its characteristics. Further disclosed are mutated proteins (X\*)  
 CC produced by the new method and with altered activity, relative to (X).  
 CC Also disclosed is a nucleic acid (I) comprising a sequence for N-  
 CC dideoxyribosyltransferase (DDTP), obtained from an N-  
 CC dideoxyribosyltransferase (DTP) by using the new method. The method is  
 CC specifically used to evolve an N-deoxyribosyltransferase (DTP) to an N-  
 CC dideoxyribosyltransferase (DDTP), i.e. an enzyme able to transfer  
 CC dideoxyribose from one dideoxyribonucleotide to another. Such an enzyme  
 CC is used for preparation of nucleotides and nucleosides that have  
 CC antitumour activity or can be used to treat infections, particularly ddc  
 CC and ddi. The current sequence represents the N-deoxyribosyltransferase  
 CC (DTP) amino acid sequence containing a G9S mutation. This protein  
 CC displays the activity of N-dideoxyribosyltransferase.

XX Sequence 157 AA;

Alignment Scores:  
 Pred. No.: 1.33e-79 Length: 157  
 Score: 842.00 Matches: 156  
 Percent Similarity: 99.36% Conservative: 0  
 Best Local Similarity: 99.36% Mismatches: 1  
 Query Match: 97.79% Indels: 0  
 DB: 8 Gaps: 0

US-10-049-750-13 (1-483) x ADR97103 (1-157)

QY 10 ATGCCAAAAGACGATCTACTTCGGTGGCGGTGGTCTCACTGACGCCCAAAACAAAGCC 69  
 Db 1 MetProLysLysThrIleTyrPheSerAlaGlyTrpPheThrAspArgGlnAsnLysAla 20  
 QY 70 TACAAGGAGCCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAACACGCTAC 129  
 Db 21 TyrLysGluAlaMetGluAlaLeuLysGluAsnProThrIleAspLeuGluAsnSerTyr 40  
 QY 130 GTTCCCTCGACCAACGATACAGGGTATCCGGTTTGATGAACACCCGGAATACCTGCAT 189  
 Db 41 ValProLeuAspAsnGlnTyrLysGlyIleArgValAspGluHisProGluTyrLeuHis 60  
 QY 190 GACAAGGTTGGGCTACGGCCACCTACAAACGACTTGAACGGGATCAAGACCAACGAC 249

Db 61 AspLysValTrpAlaThrAlaThrTyrAsnAsnAspLeuAsnGlyIleLysThrAsnAsp 80  
 QY 250 ATCATGCTGGGTCTTACATCCCTGACGAAGAAGACGTGGCTGGCGCATGGAACCTGGGT 309  
 Db 81 IleMetLeuGlyValTyrIleProAspGluGluAspValGlyLeuGlyMetGluLeuGly 100  
 QY 310 TAGCCCTTGAAGCAAGGAGTACGTCCTTTGGTTCATCCCGGACGAAGACTACGGCAAG 369  
 Db 101 TyrAlaLeuSerGlnGlyLysTyrValLeuLeuValIleProAspGluAspTyrGlyLys 120  
 QY 370 CGATCAACCTCATGAGCTGGGCGTCAGCGAACAGTGCATCAAGATGAGCCAGCTGAAG 429  
 Db 121 ProIleAsnLeuMetSerTrpGlyValSerAspAsnValIleLysMetSerGlnLeuLys 140  
 QY 430 GACTTCAACTTCAACAAAGCCGGCTTCGACTTCTACGAAGGTGCCGTATAC 480  
 Db 141 AspPheAsnPheAsnLysProArgPheAspPheTyrGluGlyAlaValTyr 157

## RESULT 3

ABP98175  
 ID ABP98175 standard; protein; 158 AA.

AC ABP98175;

XX 11-AUG-2003 (first entry)

DE Amino acid sequence of N-deoxyribosyltransferase protein.

KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;  
 KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.

XX Lactobacillus helveticus.

XX WO2003025163-A2.

XX 27-MAR-2003.

XX 12-SEP-2002; 2002WO-FR003120.

XX 14-SEP-2001; 2001FR-00011911.

XX (INSP) INST PASTEUR.

PA (NARE-) INST NAT RECH AGRONOMIQUE.

PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;

XX WPI; 2003-313351/30.

DR N-PSDB; ACC43580, ACC43586.

PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing  
 PT deoxyribonucleotides, e.g. for use as antiviral agents, also related  
 PT nucleic acid and antibodies.

XX Claim 1; Page 61; 70pp; French.

XX The present sequence represents a N-deoxyribosyltransferase protein from  
 CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-  
 CC deoxyribosyltransferase protein is useful to raise specific antibodies,  
 CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,  
 CC particularly those containing non-natural bases. These  
 CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),  
 CC parasitic or fungal infections, for antitumour chemotherapy, and as  
 CC insecticides or herbicides

XX Sequence 158 AA;

## Alignment Scores:

Pred. No.: 1,38e-67 Length: 158  
 Score: 728.00 Matches: 128  
 Percent Similarity: 95.48% Conservative: 20  
 Best Local Similarity: 82.58% Mismatches: 7  
 Query Match: 84.55% Indels: 0

DB: 6 Gaps: 0

US-10-049-750-13 (1-483) x ABP98175 (1-158)

QY 16 AAAAAGACGATCTACTTCGGTCCCGCTGGTTCACTGACCGCAAAACAAAGCCTACAAG 75  
 Db 4 LysLysThrLeuTyrPheGlyAlaGlyTrpPheAsnGluLysGlnAsnLysAlaTyrLys 23  
 QY 76 GAAGCATGGAAGCCCTCAAGGAAACCCCAAGTAGTACCTGGAAAAAGTAGTACGTTC 135  
 Db 24 GluAlaMetAlaAlaLeuLysGluAsnProThrValAspLeuGluAsnSerTyrValPro 43  
 QY 136 CTGGACACACCTACAAAGGTATCCGGTTGTATGACACCCCGGATACCTGTCATGACAAG 195  
 Db 44 LeuGluAsnGlnTyrLysGlyIleArgIleAspGluHisProGluTyrLeuHisAsnIle 63  
 QY 196 GTTTCGGCTACCGCCACCTACAAACAGACTTTGAACGGGATCAAGCAACGACATCATG 255  
 Db 64 GluTrpAlaSerAlaThrTyrHisAsnAspLeuValGlyIleLysThrSerAspValMet 83  
 QY 256 CTGGGTGTCTACATCCCTGACGAAGACGTGGCTGGGCATGGAACCTGGGTACGCC 315  
 Db 84 LeuGlyValTyrLeuProGluGluAspValGlyLeuGlyMetGluLeuGlyTyrAla 103  
 QY 316 TTGAGCCCAAGCAAGTACGTCTCTTTGGTCATCCCGGACGAAGACTACGGCAAGCCGATC 375  
 Db 104 LeuSerGlnGlyLysTyrIleLeuLeuValIleProAspGluAspTyrGlyLysProIle 123  
 QY 376 AACCTCATGAGCTGGGCGTCAGCGACACGTCGATCAAGATGAGCCAGCTGAAGGACTTTC 435  
 Db 124 AsnLeuMetSerTrpGlyValCysAspAsnAlaIleLysIleSerGluLeuLysAspPhe 143  
 QY 436 AACTTCAACAAAGCCCGCTTCGACTTCTACGAAGGTGCCGTATAC 480  
 Db 144 AspPheAsnLysProArgTyrAsnPheTyrAspGlyAlaValTyr 158

## RESULT 4

ABP98180  
 ID ABP98180 standard; protein; 159 AA.

XX AC ABP98180;

XX 11-AUG-2003 (first entry)

DE Amino acid sequence of N-deoxyribosyltransferase protein.

KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;  
 KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.

XX Lactobacillus acidophilus.

XX WO2003025163-A2.

XX 27-MAR-2003.

XX 12-SEP-2002; 2002WO-FR003120.

XX 14-SEP-2001; 2001FR-00011911.

XX (INSP) INST PASTEUR.

PA (NARE-) INST NAT RECH AGRONOMIQUE.

PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;

XX WPI; 2003-313351/30.

DR N-PSDB; ACC43585.

PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing  
 PT deoxyribonucleotides, e.g. for use as antiviral agents, also related  
 PT nucleic acid and antibodies.

XX Claim 1; Page 68; 70pp; French.









[illegible]





Db 474 ThrThrThrThrThrThrThrThrThr-----ThrThrThr 485

Qy 440 TCAACAGCCGGCTTCACCTCTACGAGGTGGGTACT 481

Db 486 ThrThrThrThrAlaThrThrThrThrThrLysProThrThr 499

RESULT 14

AB111746

ID AAB11746 standard; protein; 249 AA.

XX AAB11746;

XX 28-OCT-2000 (first entry)

XX C. parvum NINC isolate GP900 variant domain 2, SEQ ID NO:20.

XX GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;

KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;

KW merozoite; diarrhoea; protozoacide; domain 2; mucin-like; variant.

XX Cryptosporidium parvum.

OS US6071518-A.

XX US6071518-A.

XX 06-JUN-2000.

XX 12-SEP-1997; 97US-00928361.

XX 29-MAY-1992; 92US-00891301.

PR 01-JUN-1993; 93US-00071880.

PR 03-APR-1995; 95US-00415751.

PR 14-AUG-1996; 96US-00700651.

PR 13-SEP-1996; 96US-0026062P.

XX (REGC ) UNIV CALIFORNIA.

XX Petersen C;

XX WPI; 2000-422065/36.

XX New GP900 protein fragments and fusion proteins of Cryptosporidium

PT parvum useful for detecting the presence of the parasite, and diagnosing

PT or treating Cryptosporidium infections by competitive inhibition of the

PT function of GP900.

XX Claim 4; Col 17-18; 59pp; English.

XX The invention relates to the GP900 glycoprotein of the protozoan

CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion

CC proteins comprising GP900 fragments. The invention also relates to the

CC administration of GP900 or fragments thereof to a host to elicit anti-

CC GP900 antibody production, and to a method of cryptosporidiosis treatment

CC or prophylaxis comprising administration of anti-GP900 antibodies to an

CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to

CC competitively inhibit sporozoite or merozoite attachment or invasion, and

CC are also useful for the generation of anti-GP900 antibodies. The

CC antibodies also inhibit sporozoite or merozoite attachment/invasion and

CC additionally inhibit the binding of GP900 ligands to GP900. GP900

CC proteins, fragments and antibodies may therefore be used to treat or

CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common

CC cause of diarrhoea in humans and causes life-threatening diarrhoea in

CC immunocompromised persons. Cryptosporidiosis can be contracted from

CC contaminated municipal water supplies (e.g., public swimming pools). It

CC is also a cause of disease in animals, resulting in financial losses in

CC agriculture. GP900 fragments, fusion proteins and antibodies may also be

CC used for the diagnosis of Cryptosporidium parvum infections, and for the

CC detection of the parasite in the environment. The present sequence

CC represents a variant of the mucin-like domain 2 (AAB11734) of the

CC Cryptosporidium parvum NINC isolate GP900 protein (AAB11727)

XX Sequence 249 AA;

XX Alignment Scores:

Pred. No.: 2.16e-05 Length: 249

Score: 138.50 Matches: 44

Percent Similarity: 43.75% Conservative: 26

Best Local Similarity: 27.50% Mismatches: 89

Query Match: 16.09% Indels: 1

DB: 3 Gaps: 1

US-10-049-750-13 (1-483) x AAB11746 (1-249)

Qy 2 TATACTAAATGCCAAAGAGCATCTACTTCGGTGGCGGTTCATCGCCGCAAA 61

Db 9 TyrThrLysCysValGlyValLysHisThrThrThrThrThrThrThrThr 28

Qy 62 ACAAGAGCCCTACAGGAAGCCATGGAAGCCCTCAAGGAAACCCCAAGATTGACCTGGAAA 121

Db 29 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 48

Qy 122 ACAGCTAGTTCCCTCGGACAAACAGTACAAAGGTATCCGGTTCATGAACACCCCGGAAT 181

Db 49 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 68

Qy 182 ACCTGCATGACAAAGGTTTGGGCTAGGCGCACCTACACAAACGACTTGACCGGATCAAAGA 241

Db 69 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 87

Qy 242 CCAACGACATCATGCTGGGTGTCTACATCCCTCGACGAAGAAGCGTCCGCTGGGCATGG 301

Db 88 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 107

Qy 302 AACTGGGTTCAGCTTGGAGCCCAAGTACGCTTGGTTCATCCGGACGAAGACT 361

Db 108 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 127

Qy 362 ACGGCAAGCCGATCAACCTCATGAGCTGGGCGTCAGCAACACGTCATCAAGATGAGCC 421

Db 128 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 147

Qy 422 AGCTGAAGGACTTCAACTTCAACAGCCGCTTCGACTTCTACGAAGTGCCTGATCT 481

Db 148 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 167

RESULT 15

ABJ04059

ID ABJ04059 standard; protein; 249 AA.

XX AC ABJ04059;

XX 27-SEP-2002 (first entry)

XX C parvum GP900 protein fragment SEQ ID NO: 20.

XX Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis.

XX Cryptosporidium parvum.

XX WO200194631-A1.

XX 13-DEC-2001.

XX 14-MAY-2001; 2001WO-US015624.

XX 06-JUN-2000; 2000US-00588995.

XX (REGC ) UNIV CALIFORNIA.

XX Petersen C, Barnes DA, Nelson RG, Gut J;

XX WPI; 2002-566447/60.

XX Detecting Cryptosporidium in biological and environmental samples and

PT diagnosis of cryptosporidiosis involves, contacting the sample with

PT Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA.

XX

PS Disclosure; Page 128-129; 157pp; English.

XX CC The present invention relates to a method of detecting Cryptosporidium in  
CC biological and environmental samples, and of diagnosing  
CC cryptosporidiosis. This involves obtaining a sample and contacting it  
CC with Cryptosporidium GP900, P68 or cryptosporin antigen, antibody, DNA or  
CC RNA, or its variant, mutant or fragment. The method is also useful for  
CC detecting and identifying individual Cryptosporidium isolates based on  
CC the genetic characteristics, and for diagnosis of prior or concurrent  
CC Cryptosporidium infection. The present sequence is a C. parvum protein  
CC sequence used in the exemplification of the invention  
XX  
SQ Sequence 249 AA;

Alignment Scores:

Pred. No.:	2.16e-05	Length:	249
Score:	138.50	Matches:	44
Percent Similarity:	43.75%	Conservative:	26
Best Local Similarity:	27.50%	Mismatches:	89
Query Match:	16.09%	Indels:	1
DB:	5	Gaps:	1

US-10-049-750-13 (1-483) x ABJ04059 (1-249)

QY	2	TATACTAAATGCAAAAAAGACGATCTACTCGGTCCGGCTGGTTCACTGACCGCAAA	61
DB	9	TyrThrLysCysValGlyVallyHisThrThrThrThrThrThrThrThrThrThr	28
QY	62	ACAAAGCCTACAAAGGAGCCATGGAGCCCTCAAGGAAACCAACGATTGACCTGGAAA	121
DB	29	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	48
QY	122	ACAGCTACCTTCCTCCCTGGACACACAGTACAGGGTATCCGGGTGTGATGAACACCGGAAT	181
DB	49	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	68
QY	182	ACCTGCATGACAAAGTTTGGGCTACGCCACCTACAAACGACTTGAACGGGATCAAGA	241
DB	69	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	87
QY	242	CCAACGACATCATGTGGGTGTCTACATCCCTGACGAGAACGACGTCGGCCTGGCATGG	301
DB	88	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	107
QY	302	AACTGGGTTACGCTTTGAGCCCAAGGCAAGTACGTCCTTTTGGTTCATCCGGACGAAGACT	361
DB	108	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	127
QY	362	ACGGCAAGCCGATCAACCTCATGAGCTGGGGGTGACGACAAACGTGATCAAGATGAGCC	421
DB	128	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	147
QY	422	ACGTGAGGACTTCACTTCAACAGCCGGCTTCGACTTCTACGAAGTGCCTATACT	481
DB	148	ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr	167

Search completed: December 23, 2005, 22:29:56  
Job time : 96.6485 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:17:41 ; Search time 19.609 Seconds  
(without alignments)  
4072.865 Million cell updates/sec

Title: US-10-049-750-13

Perfect score: 861

Sequence: 1 gtatactaataatgccaaaaa.....acgaagggtgcgtatactaa 483

Scoring table:

BLASUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10049750/runat\_23122005\_113553\_6317/app\_query\_fasta\_1.1742  
-DB=Issued\_Patents\_AA -OFMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10049750@cgn1.1.101@runat\_23122005\_113553\_6317 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG-SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5 COMB pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6 COMB pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H COMB pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS COMB pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE COMB pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142	16.5	216	2	US-08-928-361B-8
2	142	16.5	216	2	US-08-928-361B-8
3	142	16.5	1837	2	US-08-928-361B-5
4	142	16.5	1837	2	US-08-928-361B-5
5	138.5	16.1	249	2	US-08-700-651-15
6	138.5	16.1	249	2	US-08-928-361B-20
7	138.5	16.1	249	2	US-08-928-361B-20
8	138.5	16.1	1721	2	US-08-700-651-5
9	138.5	16.1	1721	2	US-08-928-361B-6
10	138.5	16.1	1721	2	US-08-928-361B-6
11	134.5	15.6	175	2	US-08-700-651-12
12	134.5	15.6	175	2	US-08-928-361B-17

13	134.5	15.6	175	2	US-09-588-995A-17	Sequence 17, Appl
14	132	15.3	216	2	US-08-928-361B-27	Sequence 27, Appl
15	126.5	14.7	288	2	US-09-216-393B-341	Sequence 341, App
16	126.5	14.7	288	2	US-09-216-393B-344	Sequence 344, App
17	126.5	14.7	341	2	US-09-252-991A-22507	Sequence 22507, A
18	125.5	14.6	605	2	US-09-487-558B-428	Sequence 428, App
19	123	14.3	162	2	US-08-700-651-13	Sequence 13, Appl
20	120.5	14.0	618	2	US-09-252-991A-28358	Sequence 28358, A
21	119.5	13.9	618	2	US-09-252-991A-23696	Sequence 23696, A
22	117	13.6	1113	2	US-09-252-991A-29215	Sequence 29215, A
23	113	13.1	138	2	US-08-700-651-10	Sequence 10, Appl
24	113	13.1	138	2	US-08-928-361B-15	Sequence 15, Appl
25	113	13.1	138	2	US-09-588-995A-15	Sequence 15, Appl
26	113	13.1	181	2	US-09-252-991A-21966	Sequence 21966, A
27	112.5	13.1	150	2	US-08-928-361B-18	Sequence 18, Appl
28	112.5	13.1	150	2	US-09-588-995A-18	Sequence 18, Appl
29	112.5	13.1	658	2	US-09-252-991A-25861	Sequence 25861, A
30	111.5	13.0	611	2	US-09-252-991A-17597	Sequence 17597, A
31	110.5	12.8	406	2	US-09-252-991A-19857	Sequence 19857, A
32	109.5	12.7	508	2	US-09-252-991A-25549	Sequence 25549, A
33	109	12.7	130	2	US-08-700-651-8	Sequence 8, Appl
34	109	12.7	130	2	US-08-928-361B-13	Sequence 13, Appl
35	109	12.7	130	2	US-09-588-995A-13	Sequence 13, Appl
36	108.5	12.6	154	2	US-09-252-991A-32553	Sequence 32553, A
37	108.5	12.6	174	2	US-09-252-991A-27275	Sequence 27275, A
38	108.5	12.6	2870	2	US-09-479-467A-15	Sequence 15, Appl
39	108.5	12.6	2870	2	US-09-655-160-15	Sequence 15, Appl
40	108.5	12.6	3178	2	US-09-479-467A-4	Sequence 4, Appl
41	108.5	12.6	3178	2	US-09-655-160-4	Sequence 4, Appl
42	108	12.5	130	2	US-08-700-651-9	Sequence 9, Appl
43	108	12.5	130	2	US-08-928-361B-14	Sequence 14, Appl
44	108	12.5	130	2	US-09-588-995A-14	Sequence 14, Appl
45	108	12.5	453	2	US-09-686-583B-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-08-928-361B-8  
; Sequence 8, Application US/08928361B  
; Patent No. 6071518

; GENERAL INFORMATION:

; APPLICANT: Petersen, Carolyn

; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,

; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM

; TITLE OF INVENTION: SPECIES INFECTIONS

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PETERS, VERNY, JONES & BIKSA

; STREET: 385 Sherman Avenue, Suite 6

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94306-1840

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/928,361B

; FILING DATE: 12-SEP-1997

; CLASSIFICATION:

; PRIOR APPLICATION DATA: US 60/026,062

; APPLICATION NUMBER: 13-SEP-1996

; FILING DATE: 13-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: VERNY, Hana

; REGISTRATION NUMBER: 30,518

; REFERENCE/DOCKET NUMBER: 480.76-1 (HV)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-324-1677

STATE: CA  
COUNTRY: USA  
ZIP: 94306-1840  
COMPUTER READABLE  
MEDIUM TYPE: FLOPPY  
COMPUTER: IBM PC



Alignment Scores:		
Pred. No.:	4, 79e-07	1837
Score:	142.00	45
Percent Similarity:	46.10%	Conservative: 26
Best Local Similarity:	29.22%	Mismatches: 67
Query Match:	16.49%	Indels: 16
DB:	2	Gaps: 3

US-10-049-750-13 (1-483) x US-09-588-995A-5 (1-1837)

[illegible]

## RESULT 5

US-08-700-651-15  
; Sequence 15, Application US/08700651B  
; Patent No. 6015882  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN

APPLICANT: LEECH, JAMES  
APPLICANT: NELSON, RICHARD, C.  
APPLICANT: GUT, JIRI  
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
TITLE OF INVENTION: INFECTIONS  
FILE REFERENCE: 480.19-4(HV)  
CURRENT APPLICATION NUMBER: US/08/700,651B  
CURRENT FILING DATE: 1997-08-14  
EARLIER APPLICATION NUMBER: 08/415,751  
EARLIER FILING DATE: 1995-04-03  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 15  
LENGTH: 249  
TYPE: PRT  
ORGANISM: Cryptosporidium parvum  
FEATURE:  
OTHER INFORMATION: mutant/variant of SEQ ID NO:5  
US-08-700-651-15

Alignment Scores:  
Pred. No.: 5,228-07 Length: 249  
Score: 138.50 Matches: 44  
Percent Similarity: 43.75% Conservative: 26  
Best Local Similarity: 27.50% Mismatches: 89  
Query Match: 16.09% Indels: 1  
DB: 2 Gaps: 1

US-10-049-750-13 (1-483) x US-08-700-651-15 (1-249)

```
QY 2 TATACATAAATGCAAAAAGACGATCTACTTCGGTCCGGCTGGTTCACTGACCGCCAAA 61
Db 9 TyrThrLysCysValGlyValLysHisThrThrThrThrThrThrThrThrThrThrThr 28
QY 62 ACAAGCCTACAAAGGAGCCATGGAAGCCCTCAAGGAAACCCACGATTGACCTGGAAA 121
Db 29 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 48
QY 122 ACAGCTAGCTTCCCTGGACACCAAGGATACAGGGTATCCGGTGTGATGAACACCGGAAT 181
Db 49 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 68
QY 182 ACCTGCATGACAAAGTTGGGCTAGCGCCACTACAAACGACTTGACGGGATCAAGA 241
Db 69 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 87
QY 242 CCAAGCATCATGCTGGGTGTCTACATCCCTGACGAAAGACGTCGGCCTGGGCATGG 301
Db 88 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 107
QY 302 AACTGGGTTACGCCCTTGAGCCAAAGCAAGTACGTCCTTTGGTCAATCCCGGACGAGACT 361
Db 108 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 127
QY 362 ACGCAAGCCGATCAACTCATGAGCTGGGGCTGAGCAACAGTGTCAAGATGAGCC 421
Db 128 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 147
QY 422 ACCTGAGGACTTCACCTCAACAGCCGGCTTCGACTTCTACGAAGTGCCTGATACT 481
Db 148 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 167
```

RESULT 6

US-08-928-361B-20  
Sequence 20, Application US/08928361B  
Patent No. 6071518  
GENERAL INFORMATION:  
APPLICANT: Petersen, Carolyn  
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
TITLE OF INVENTION: SPECIES INFECTIONS

NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-1840  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,361B  
FILING DATE: 12-SEP-1997

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,062  
FILING DATE: 13-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Verny, Hana  
REGISTRATION NUMBER: 30,518  
REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-324-1677  
TELEFAX: 650-324-1678  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 249 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-928-361B-20

Alignment Scores:  
Pred. No.: 5,228-07 Length: 249  
Score: 138.50 Matches: 44  
Percent Similarity: 43.75% Conservative: 26  
Best Local Similarity: 27.50% Mismatches: 89  
Query Match: 16.09% Indels: 1  
DB: 2 Gaps: 1

US-10-049-750-13 (1-483) x US-08-928-361B-20 (1-249)

```
QY 2 TATACATAAATGCAAAAAGACGATCTACTTCGGTCCGGCTGGTTCACTGACCGCCAAA 61
Db 9 TyrThrLysCysValGlyValLysHisThrThrThrThrThrThrThrThrThrThrThr 28
QY 62 ACAAGCCTACAAAGGAGCCATGGAAGCCCTCAAGGAAACCCACGATTGACCTGGAAA 121
Db 29 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 48
QY 122 ACAGCTAGCTTCCCTGGACACCAAGGATACAGGGTATCCGGTGTGATGAACACCGGAAT 181
Db 49 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 68
QY 182 ACCTGCATGACAAAGTTGGGCTAGCGCCACTACAAACGACTTGACGGGATCAAGA 241
Db 69 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 87
QY 242 CCAAGCATCATGCTGGGTGTCTACATCCCTGACGAAAGACGTCGGCCTGGGCATGG 301
Db 88 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 107
QY 302 AACTGGGTTACGCCCTTGAGCCAAAGCAAGTACGTCCTTTGGTCAATCCCGGACGAGACT 361
Db 108 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 127
QY 362 ACGCAAGCCGATCAACTCATGAGCTGGGGCTGAGCAACAGTGTCAAGATGAGCC 421
Db 128 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 147
```



[illegible]

[illegible][illegible]

**THIS PAGE BLANK (USPTO)**

Result No.	Query			ID	Description
	Score	Match	Length		
1	126.5	14.7	288	US-09-216-393-341	Sequence 3
2	126.5	14.7	288	US-09-216-393-344	Sequence 3
3	126.5	14.7	288	US-10-321-856-341	Sequence 3
4	126.5	14.7	288	US-10-321-856-344	Sequence 3
5	126	14.6	1795	US-11-037-143-36210	Sequence 3
6	125.5	14.6	605	US-09-801-368-428	Sequence 4
7	125.5	14.6	605	US-10-369-493-22016	Sequence 2
8	124	14.4	292	US-10-425-114-59190	Sequence 5
9	119.5	13.9	19723	US-10-084-846A-5	Sequence 5
10	119	13.8	324	US-10-425-114-53603	Sequence 5
11	119	13.8	347	US-10-425-114-56484	Sequence 5

Db 139 ArgSerLysArgGlyLysThrThrThrThrSerSerSerThrSerThrSerThr 158

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```
QY 68 CTTACAGGAGCCATGGAAGCCCTCAAGAAACCAACCAAGATTGACCTGGAAAAACAGCT 127
Db 159 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
QY 128 ACCTTCCCTGGGCAACCCAGTACAAAGGGTATCGGGTTGATGAACACCCGGATACCTGC 187
Db 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThrThrThrThrThr 198
QY 188 ATGACACAGGTTGGGCTACCGCCACCTCAACAACGACTTGAACGGGATCAAGACCAACG 247
Db 199 ThrThr-----ThrThrProThrThrThrThr----- 215
QY 248 ACATCATGCTGGGTGCTCATCCCTGACGAAAGACGTCGGCCTGGGCATGGAACCTGG 307
Db 216 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrPro 241
QY 308 GTTACGCCCTTGGAGCAAGCAAGTACGCTCTTTTGGTCATCCCGGACGAAGACTACGGCA 367
Db 226 -----ThrThrProThr----- 225
QY 368 AGCGATCAACCTCATGAGCTGGGGCTCAGCGACAACAGTGATCAAGATGAGCCAGCTGA 427
Db 242 ThrThrThrThrThrThrThrThrGluProThrThrThrThrThrThrGluProThrThr 261
QY 428 AGGACTTCAACTTCAACAAGCCGCGCTTCGACTTCTACGAAGGTGCCGTATACT 481
Db 262 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrProSerThr 279

RESULT 2
US-09-216-393-344
; Sequence 344, Application US/09216393
; Patent No. US2001001447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; EARLIER FILING DATE: 1998-12-18
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 344
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-09-216-393-344

Alignment Scores:
Pred. No.: 0.000556 Length: 288
Score: 126.50 Matches: 41
Percent Similarity: 43.67% Conservative: 28
Best Local Similarity: 25.95% Mismatches: 72
Query Match: 14.69% Indels: 17
DB: 3 Gaps: 2

US-10-049-750-13 (1-483) x US-09-216-393-344 (1-288)
QY 8 AAATGCCAAAAGACAGATCTACTTCGGTCCGCGCTGGTTCAGTACCGCCAAAACAAAG 67
Db 139 ArgSerLysArgGlyLysLysThrThrThrThrThrThrThrSerSerThrThrThrThr 158
QY 68 CCTACAGGAAGCCATGGAAGCCCTCAAGAAACCAACCAAGATTGACCTGGAAACAGCT 127
Db 159 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
QY 128 ACCTTCCCTGGCAACACCAAGTATCCGGTTCAGTACCGCCGAAAGCAAG 187
Db 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThrThrThrThrThr 198
QY 188 ATGACAGGTTGGGCTACCGCCACCTCAAGAAACCAACCAAGTTCGAGTCAAGACCAAGCT 247
Db 199 ThrThr-----ThrThrProThrThrThrThr----- 215
QY 248 ACCTTCCCTGGCAACACCAAGTATCCGGTTCAGTACCGCCGAAAGCAAG 307
Db 216 ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThrThrThrThrThr 198
QY 188 ATGACAGGTTGGGCTACCGCCACCTCAAGAAACCAACCAAGTTCGAGTCAAGACCAAGCT 247
Db 226 -----ThrThrProThr----- 225
QY 308 GTTACGCCCTTGGAGCAAGCAAGTACGCTCTTTTGGTCATCCCGGACGAAGACTACGGCA 367
Db 226 -----ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrProThr 241
```

```
Db 199 ThrThr-----ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThr 215
QY 248 ACATCATGCTGGGTGCTCATCCCTGACGAAAGACGTCGGCCTGGGCATGGAACCTGG 307
Db 216 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrPro 225
QY 308 GTTACGCCCTTGGAGCAAGCAAGTACGCTCTTTTGGTCATCCCGGACGAAGACTACGGCA 367
Db 226 -----ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrProThr 241
QY 368 AGCGATCAACCTCATGAGCTGGGGCTCAGCGACAACAGTGATCAAGATGAGCCAGCTGA 427
Db 242 ThrThrThrThrThrThrThrThrGluProThrThrThrThrThrThrGluProThrThr 261
QY 428 AGGACTTCAACTTCAACAAGCCGCGCTTCGACTTCTACGAAGGTGCCGTATACT 481
Db 262 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrProSerThr 279

RESULT 3
US-10-321-856-341
; Sequence 341, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THERE
; FILE REFERENCE: TX-1-C2-1
; CURRENT APPLICATION NUMBER: US/10/321,856
; CURRENT FILING DATE: 2002-12-17
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 09/216,393
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 341
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-10-321-856-341

Alignment Scores:
Pred. No.: 0.000556 Length: 288
Score: 126.50 Matches: 41
Percent Similarity: 43.67% Conservative: 28
Best Local Similarity: 25.95% Mismatches: 72
Query Match: 14.69% Indels: 17
DB: 4 Gaps: 2

US-10-049-750-13 (1-483) x US-10-321-856-341 (1-288)
QY 8 AAATGCCAAAAGACAGATCTACTTCGGTCCGCGCTGGTTCAGTACCGCCAAAACAAAG 67
Db 139 ArgSerLysArgGlyLysLysThrThrThrThrThrThrThrSerSerThrThrThrThr 158
QY 68 CCTACAGGAAGCCATGGAAGCCCTCAAGAAACCAACCAAGATTGACCTGGAAACAGCT 127
Db 159 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
QY 128 ACCTTCCCTGGCAACACCAAGTATCCGGTTCAGTACCGCCGAAATACCTGC 187
Db 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrProThrThrThrThrThr 198
QY 188 ATGACAGGTTGGGCTACCGCCACCTCAACAACGACTTGAACGGGATCAAGACCAACG 247
Db 199 ThrThr-----ThrThrProThrThrThrThr----- 215
QY 248 ACATCATGCTGGGTGCTCATCCCTGACGAAAGACGTCGGCCTGGGCATGGAACCTGG 307
Db 216 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrPro 225
QY 308 GTTACGCCCTTGGAGCAAGCAAGTACGCTCTTTTGGTCATCCCGGACGAAGACTACGGCA 367
Db 226 -----ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrProThr 241
```



```
QY 368 AGCGATCAACTCATGAGTGGGCGTCAAGCAACAGTGCATCAAGATGAGCGAGTGA 427
   ::::::::::: ::::: ||||| :::::
Db 242 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 261
   ||||| ||||| ::::: ||||| :::::
QY 428 AGGACTTCAACTCAACAGCGCGCTTCGACTTCTACGAGGTGCGGTATACT 481
   ||||| ||||| ::::: ||||| :::::
Db 262 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 279

RESULT 4
US-10-321-856-344
; Sequence 344, Application US/10321856
; Publication No. US20030194393A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THEREOF
; FILE REFERENCE: TX-1-C2-1
; CURRENT APPLICATION NUMBER: US/10/321,856
; CURRENT FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 09/216,393
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/994,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 344
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-10-321-856-344

Alignment Scores:
Pred. No.: 0.000556 Length: 288
Score: 126.50 Matches: 41
Percent Similarity: 43.67% Conservative: 28
Best Local Similarity: 25.95% Mismatches: 72
Query Match: 14.63% Indels: 17
DB: 4 Gaps: 2

US-10-049-750-13 (1-483) x US-10-321-856-344 (1-288)
QY 8 AAATGCCAAAAGACGATCTACTTCGGTGGCGGTCTACTGACCGCAAAACAAG 67
   ::::::::::: ::::: ||||| :::::
Db 139 ArgSerLysArgGlyLysThrThrThrThrThrThrThrThrThrThrThrThrThr 158
   ||||| ||||| ::::: ||||| :::::
QY 68 CTTCAAGAGAGCCATGGAAGCCCTCAAGGAAACCAAGCATGACCTGGGAAACAGCT 127
   ||||| ||||| ::::: ||||| :::::
Db 159 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 178
   ||||| ||||| ::::: ||||| :::::
QY 128 AGGTTCCCTGGACCAACGATCAAGGGTATCCGGGTTGATGAACACCGGAAATACCTGC 187
   ||||| ||||| ::::: ||||| :::::
Db 179 ThrThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 198
   ||||| ||||| ::::: ||||| :::::
QY 188 ATGACAAAGTTGGGTACGGCCACTCAACAAAGCATGTAACGGGATCAAGCAACG 247
   ||||| ||||| ::::: ||||| :::::
Db 199 ThrThr-----ThrThrProThrThrThrThrThrThrThrThrThrThrThr 215
   ||||| ||||| ::::: ||||| :::::
QY 248 ACATCATGCTGGGTCTACATCCCTGACGAGAAGAGCGTCGGCTGGCGATGGAACG 307
   ||||| ||||| ::::: ||||| :::::
Db 216 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 225
   ||||| ||||| ::::: ||||| :::::
QY 308 GTTACGCTTGAAGCAAGTACGTCCTTTTGGTTCATCCGCGACGAAAGTACGGCA 367
   ::::::::::: ::::: ||||| :::::
Db 226 -----ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 241
   ||||| ||||| ::::: ||||| :::::
QY 368 AGCGATCAACTCATGAGTGGGCGTCAAGCAACAGTGCATCAAGATGAGCGAGTGA 427
   ::::::::::: ::::: ||||| :::::
Db 242 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 261
   ||||| ||||| ::::: ||||| :::::
QY 428 AGGACTTCAACTCAACAGCGCGCTTCGACTTCTACGAGGTGCGGTATACT 481
   ||||| ||||| ::::: ||||| :::::
Db 262 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 279
   ||||| ||||| ::::: ||||| :::::
```

## RESULT 5

```
US-11-097-143-36210
; Sequence 36210, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36210
; LENGTH: 1795
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-36210

Alignment Scores:
Pred. No.: 0.00107 Length: 1795
Score: 126.00 Matches: 50
Percent Similarity: 47.02% Conservative: 29
Best Local Similarity: 29.76% Mismatches: 73
Query Match: 14.63% Indels: 16
DB: 6 Gaps: 5

US-10-049-750-13 (1-483) x US-11-097-143-36210 (1-1795)
QY 5 ACTAATGCCAAAAGACGATCTACTTCGGTGGCGGTCTACTGACCGCAAAACA 64
   ||||| ::::: ||||| :::::
Db 493 ThrLysSerThrProLysIleSerSerThrThrGluGlnHisSerThrThrThrAlaLysThr 512
   ||||| ||||| ::::: ||||| :::::
QY 65 AAGCCTCAAGAGAGCCATGGAAGCCCTCAAGGAAACCAAGCATGACCTGGGAAACA 124
   ||||| ||||| ::::: ||||| :::::
Db 513 ThrThrThrLysArgProThrThrValThrGluLysThrSerSerAlaThrGluLysPro 532
   ||||| ||||| ::::: ||||| :::::
QY 125 GCTAGTTCCCTCGACCAACCAAGTACA-----AGGTATCCGGGTTCATGAACACC--- 175
   ||||| ||||| ::::: ||||| :::::
Db 533 ArgThrThrValThrThrThrThrGlnLysArgSerThrThrHisAsnThrSer 552
   ||||| ||||| ::::: ||||| :::::
QY 176 -----CGAATACCTGCATCAAGAGTTGGGCTACGCCACCTACCAACAGACTTGA 229
   ||||| ||||| ::::: ||||| :::::
Db 553 ProAspThrLysThrIleArgSerThrThrLeuSerProLysThrThrThrThrPro 572
   ||||| ||||| ::::: ||||| :::::
QY 230 ACGGATCAAGACCA-----ACGACATCATGCTGGGTGTCTACATCCCTGACGAAGAAGACG 286
   ||||| ||||| ::::: ||||| :::::
Db 573 SerThrThrProSerThrThrThrProSerThrThrThrProSerThrThrThrPro 592
   ||||| ||||| ::::: ||||| :::::
QY 287 TCGGCTGGGATGGAAGTACGCTTACGCTTGAAGCAAGCAAGTACGCTCTTTTGGTCA 346
   ||||| ||||| ::::: ||||| :::::
Db 593 SerThrThrThrProSerThrThrThrProSerThrThrThrValLysValSerThr 612
   ||||| ||||| ::::: ||||| :::::
QY 347 TCCCGGAGAGAGACTACGGGAGCGGATCAACCTCATGAGCTGGGGCGTACCGGACGACG 406
   ||||| ||||| ::::: ||||| :::::
```

Db 613 HisArgProArgThrThrSerGlnLysThrThr-----AlaSerThrThrThr 629  
QY 407 TGATCAAGATGAGCCAGCTGAAGACT-----TCAACTTCAACA 445  
Db 630 LysLysThrThrThrSerProLysThrThrLysThrThrAspIleProThrSerThrThr 649  
QY 446 AGCGCGCTTCGACTTCTACGAAG 469  
Db 650 SerLysLeuSerThrThrThrGln 657

RESULT 6  
US-09-801-368-428  
; Sequence 428, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 428  
; LENGTH: 605  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-428

Alignment Scores:  
Pred. No.: 0.000878 Length: 605  
Score: 125.50 Matches: 51  
Percent Similarity: 45.96% Conservative: 23  
Best Local Similarity: 31.68% Mismatches: 48  
Query Match: 14.58% Indels: 39  
DB: 3 Gaps: 7

US-10-049-750-13 (1-483) x US-09-801-368-428 (1-605)

QY 5 ACTAAATGCCAAAAGACGATCTACTTCGGTCCGGCTGGTTCACTGACGCCAACA 64  
Db 181 ThrThrThrSerThrLysLeuSerThrSerIleProThr---SerThrThrSerThr 199  
QY 65 AAGCTACAGGAGCCATGGAAGCCCTCAAGGAAACCCAAACGATTGACCTGGAACA 124  
Db 200 SerThrThrThrSerThrSerSerThrThrThrValSerValThrSerSerThr 219  
QY 125 GCTACGTTCCCTCGACACAGGATATCCGGTGTGATGAACACCCGGAATACC 184  
Db 220 SerThr-----ThrThrSerThrThrSerThrLeuIleSerThr----- 233  
QY 185 TGCATGACAGGTTGGGCTACGGCCACTTACAAACAGACTTGAACGGGATCAAGACCA 244  
Db 234 -----SerThrSerSerSerSerSerSerSerThrPro 243  
QY 245 ACGACA-----TCATGCTGGGTGTCTACATCCCTGACGAAGAGACGTCGGCTGGCA 298  
Db 244 ThrThrThrSerSerAlaProIleSerThrThrThrSerSerSerThr----- 261

US-10-049-750-13 (1-483) x US-10-369-493-22016 (1-605)

QY 5 ACTAAATGCCAAAAGACGATCTACTTCGGTCCGGCTGGTTCACTGACGCCAACA 64  
Db 181 ThrThrThrSerThrLysLeuSerThrSerIleProThr---SerThrThrSerThr 199  
QY 65 AAGCTACAGGAGCCATGGAAGCCCTCAAGGAAACCCAAACGATTGACCTGGAACA 124  
Db 200 SerThrThrThrSerThrSerSerThrThrThrValSerValThrSerSerThr 219  
QY 125 GCTACGTTCCCTCGACACAGGATATCCGGTGTGATGAACACCCGGAATACC 184  
Db 220 SerThr-----ThrThrSerThrThrSerThrLeuIleSerThr----- 233  
QY 185 TGCATGACAGGTTGGGCTACGGCCACTTACAAACAGACTTGAACGGGATCAAGACCA 244  
Db 234 -----SerThrSerSerSerSerSerSerSerThrPro 243  
QY 245 ACGACA-----TCATGCTGGGTGTCTACATCCCTGACGAAGAGACGTCGGCTGGCA 298  
Db 244 ThrThrThrSerSerAlaProIleSerThrThrThrSerSerSerThr----- 261

QY 299 TGGAACTGGGTACGCTTACGCCAAGGCAAGTACGTCCTTTTGGTCAATCCCGGACGAAG 358  
Db 262 -----SerThrSerThrThrSerProThrSerSerSerAlaPro 274  
QY 359 ACTACGGCAAGCGCATCAACCTCATGAGCTGGGGCGCTCAGCGACAACGTGATCAAGATGA 418  
Db 275 ThrSerSerSerAsnThrThr-----ProThrSerThrThr----- 286  
QY 419 GCAGCTGAAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGGTGCGGTAT 478  
Db 287 -----PheThrThrThrSerProSerThrAlaProSerSerThrThrValThrTyr 303

QY 479 ACT 481  
Db 304 Thr 304

RESULT 7  
US-10-369-493-22016  
; Sequence 22016, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 22016  
; LENGTH: 605  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22016

Alignment Scores:  
Pred. No.: 0.000878 Length: 605  
Score: 125.50 Matches: 51  
Percent Similarity: 45.96% Conservative: 23  
Best Local Similarity: 31.68% Mismatches: 48  
Query Match: 14.58% Indels: 39  
DB: 4 Gaps: 7

US-10-049-750-13 (1-483) x US-10-369-493-22016 (1-605)

QY 5 ACTAAATGCCAAAAGACGATCTACTTCGGTCCGGCTGGTTCACTGACGCCAACA 64  
Db 181 ThrThrThrSerThrLysLeuSerThrSerIleProThr---SerThrThrSerThr 199  
QY 65 AAGCTACAGGAGCCATGGAAGCCCTCAAGGAAACCCAAACGATTGACCTGGAACA 124  
Db 200 SerThrThrThrSerThrSerSerThrThrThrValSerValThrSerSerThr 219  
QY 125 GCTACGTTCCCTCGACACAGGATATCCGGTGTGATGAACACCCGGAATACC 184  
Db 220 SerThr-----ThrThrSerThrThrSerThrLeuIleSerThr----- 233  
QY 185 TGCATGACAGGTTGGGCTACGGCCACTTACAAACAGACTTGAACGGGATCAAGACCA 244  
Db 234 -----SerThrSerSerSerSerSerSerSerThrPro 243  
QY 245 ACGACA-----TCATGCTGGGTGTCTACATCCCTGACGAAGAGACGTCGGCTGGCA 298  
Db 244 ThrThrThrSerSerAlaProIleSerThrThrThrSerSerSerThr----- 261

US-10-049-750-13 (1-483) x US-10-369-493-22016 (1-605)

QY 5 ACTAAATGCCAAAAGACGATCTACTTCGGTCCGGCTGGTTCACTGACGCCAACA 64  
Db 181 ThrThrThrSerThrLysLeuSerThrSerIleProThr---SerThrThrSerThr 199  
QY 65 AAGCTACAGGAGCCATGGAAGCCCTCAAGGAAACCCAAACGATTGACCTGGAACA 124  
Db 200 SerThrThrThrSerThrSerSerThrThrThrValSerValThrSerSerThr 219  
QY 125 GCTACGTTCCCTCGACACAGGATATCCGGTGTGATGAACACCCGGAATACC 184  
Db 220 SerThr-----ThrThrSerThrThrSerThrLeuIleSerThr----- 233  
QY 185 TGCATGACAGGTTGGGCTACGGCCACTTACAAACAGACTTGAACGGGATCAAGACCA 244  
Db 234 -----SerThrSerSerSerSerSerSerSerThrPro 243  
QY 245 ACGACA-----TCATGCTGGGTGTCTACATCCCTGACGAAGAGACGTCGGCTGGCA 298  
Db 244 ThrThrThrSerSerAlaProIleSerThrThrThrSerSerSerThr----- 261

Db 262 -----SerThrSerThrThrThrSerProThrSerSerSerAlaPro 274  
QY 359 ACTACGGCAAGCCGATCAACCTCATGAGCTGGGGCGTCAGCGACAACGTGATCAAGATGA 418  
Db 275 ThrSerSerSerAsnThrThr-----ProThrSerThrThr----- 286  
QY 419 GCCAGCTGAAGGACTTCAACTTCAACAAGCGCGCTTCGACTTCTACGAAGGTGCGGTAT 478  
Db 287 -----PheThrThrThrSerProSerThrAlaProSerSerThrThrValThrTyr 303  
QY 479 ACT 481  
Db 304 Thr 304  
RESULT 8  
US-10-425-114-59190  
; Sequence 59190, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; PRIOR FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 59190  
; LENGTH: 292  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700377772\_FLI pep  
US-10-425-114-59190

## Alignment Scores:

Pred. No.: 0.00101 Length: 292  
Score: 124.00 Matches: 54  
Percent Similarity: 42.5% Conservative: 26  
Best Local Similarity: 28.7% Mismatches: 62  
Query Match: 14.4% Indels: 46  
DB: 4 Gaps: 11

US-10-049-750-13 (1-483) x US-10-425-114-59190 (1-292)

QY 14 CAAAGAACGATCTACTTTCGGTGGCGGTGT-----TCACGTGACCGCCAAACA 64  
Db 21 GlnGluArgThrSerThrThrGlnProSerSerThrProThrArgCysAlaArgAla 40  
QY 65 AGCCCTACAAGG-----AAGCATCGAAGCCC-----TCAGGAAACCCACGANTG 112  
Db 41 ArgProSerArgProLeuAspProThrArgGlyCysTyrArgArgProArgGlnArgPro 60  
QY 113 ACCTGGAAACAGTACGTTCCTCCGACACAGTACAAAGGGTATCCGG-----TTG 166  
Db 61 SerTppProProSer-----TrpTrpArgAlaCysArgAlaProGlyArgGlu 77  
QY 167 ATGAACACCCGAATACCTGCATGACAAAGTTTGGGCTA----- 205  
Db 78 ProArgArgArgThrThrArgArgThrGlyThrThrProLeuAlaAlaMet 97  
QY 206 -----CGCCACCTACACACGACTTGAACGGGA 235  
Db 98 GlyGlySerSerThrProThrProThrProThrProGlyThrThrProThrGlyProPro 117  
QY 236 TCAAGACCAACGACATCATGCTGGGTGTCTACATCCTGACGAAGACGTCGCGCTGG 295  
Db 118 AlaArgProSerThrSer-----ValThrThrSerPheSerArgArgThrThrArg 135

QY 296 GCATGG-----AACTGGGTTCAGCCCTTGAGCCAAGCA 328  
Db 136 ArgTrpCysArgProProThrProArgHisSerGlySerAlaThrPro-----Ala 152  
QY 329 AGTAGCTCCTTTTGGTTCATCCCGGACGAGACGATACGCAAGCCGATCAACCTCATGAGCT 388  
Db 153 ThrThr-----TrpArgArgArgProSerThrAlaAlaAlaAlaGlyAlaAla 170  
QY 389 GGGCGGTACGACACAGTGTATCAAGATGAGCCAGC-----TGAAGGACTTCAACTTCAACA 445  
Db 171 AlaSerSerArgThrAlaProSerArgTrpArgSerProProArgAlaProThrThrSer 190  
QY 446 AGCCGCGCTTCGACTTCTACGAAG 469  
Db 191 SerProArgProThrGlyAlaArg 198

## RESULT 9

US-10-084-846A-5  
; Sequence 5, Application US/10084846A  
; Publication No. US20040006026A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUHLNBERG, AGNES  
; APPLICANT: TREFFER, AXEL  
; APPLICANT: BECHTOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; PRIOR FILING DATE: 2003-02-25  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR FILING DATE: 2001-09-16  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patent In Ver. 3.2  
; SEQ ID NO 5  
; LENGTH: 19723  
; TYPE: PRT  
; ORGANISM: Streptomyces viridochromogenes  
; FEATURE:  
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.  
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.

## US-10-084-846A-5

Alignment Scores:  
Pred. No.: 0.0102 Length: 19723  
Score: 119.50 Matches: 54  
Percent Similarity: 28.1% Conservative: 16  
Best Local Similarity: 21.6% Mismatches: 62  
Query Match: 13.8% Indels: 117  
DB: 4 Gaps: 9

US-10-049-750-13 (1-483) x US-10-084-846A-5 (1-19723)

QY 11 TCCCAAAAAGACGATCTACTTCGGTGGCGGTGTCTACGTGACCGCCAAACAAGCT 70  
Db 9192 CysArgProAlaTrpThrThrCysProAlaGlySerArgThrSerThrArgSerPro 9211  
QY 71 ACAAGGAAG----- 79  
Db 9212 ThrArgSerGlySerArgCysGlyProThrProSerArgSerProSerArgGlyArg 9231  
QY 80 -----CCATGG----- 85  
Db 9232 SerThrThrSerSerThrTrpArgArgProProProTrpThrIleTrpArgCysArg 9251  
QY 86 -----AAGCCCTCAAGGAAACCCAA---CGATTGACCTGGAAA 121  
Db 9252 TrpArgArgCysGlySerAlaArgProGlyProArgThrArgCysGlyTrpArgTrpArg 9271  
QY 122 ACAGCTACGTTCCCTCG-----ACA 142  
Db 9272 ThrAlaProAlaSerTrpSerProProArgArgSerThrGlyIleProProSerThr 9291

```
QY 143 ACCAGTACAGGGTATCCGGTTGATGAACACCCGGATACC-----TGCATGACA 193
Db 9292 ArgSerProAlaThrGlyAlaThrSerThrArgSerAlaArgValArgCysThrThr 9311
QY 194 AGG-----TTTGGCTACGGCCACCT----- 214
Db 9312 ArgProAlaGlySerProAlaArgProProThrArgAlaArgTTPAlaArgThrPro 9331
QY 215 -----ACAACAACGACTTGAACGGGATCAAGACCAACGACATCATCTGCG 259
Db 9332 AlaLeuProAlaCysSerThrProThrAspArgGlyCysAlaAlaThrThrGlyGlyTip 9351
QY 260 GTGTCTACATCCCTGACGAG----- 280
Db 9352 CysArgProSerSerThrArgArgSerProAspCysArgSerProSerThrAlaAlaAla 9371
QY 281 -----AAGACGTCCGGCTGGGCATCGAAGTGGTTACG 313
Db 9372 ProArgProGlyArgCysAlaThrSerArgThrArgCysGlyAspTTPProThrArg 9391
QY 314 CTTGAGCCCAAGGCAAGTACGCTCTTT----- 340
Db 9392 ProSerArgAlaArgSerThrSerAlaArgProValArgProSerValArgSerArgArg 9411
QY 341 -----TGTCATCCCGGACGAGACTACGGCAAGCCGATCAACC 379
Db 9412 ArgSerProHisSerArgGluTTPSerCysGlyProSerSerAlaArgArgArgThr 9431
QY 380 TCATGAGCTGGGGCGTCAGCGACAACG 406
Db 9432 SerProAlaGlyAlaAlaProThrSer 9440
```

## RESULT 10

```
US-10-425-114-53603
; Sequence 53603, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53603
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17202D01_FLI.pep
US-10-425-114-53603
```

```
Alignment Scores:
Pred. No.: 0.00345 Length: 324
Score: 119.00 Matches: 48
Percent Similarity: 42.47% Conservative: 14
Best Local Similarity: 32.88% Mismatches: 44
Query Match: 13.82% Indels: 40
DB: 4 Gaps: 9
```

US-10-049-750-13 (1-483) x US-10-425-114-53603 (1-324)

```
QY 75 GGAAGCCATGAAGCCCTCAAGAAAACCCCAAGTTCGATGACCTGGAACACAG----- 125
Db 155 GlyLeuHisGlyGlyArgArgAlaLeu-----ProGlyArgHisArgArgAla 170
QY 126 -----CTAGTTCCTCCCTGGACACCAAGGTATCCG 161
```

```
Db 171 AlaAlaArgGlnArgValProLeuLeuArgValProGlyGlnProAlaArgHisGlnPro 190
QY 162 GGTGTGATGAACACCCGGAATACCTGCATGACAAAGGTTTGGGTACGCGCAC----- 212
Db 191 Gly-----LeuArgHisValProAla-----GlyHisHisArgGln 202
QY 213 ---CTACAACAACGACTTGAACGGGATCAAGACCAACGACATCATCTGGTGTCTACAT 269
Db 203 GlyGlnArgGlnArgProGlnLeuGlnProValArgArgHisGlyGly----- 219
QY 270 CCCTGACGAAGAAGACGTGGCGCTGGCATGGTGAACCTGGTTCGCTTGAAGCCCAAGSCAA 329
Db 220 -----ArgArgArgArgArgAlaArgGluGlyArgAlaGlnArgGlnArgArg 237
QY 330 GTAGCTCCTTTTGGTCATCCCGGACGAAGACTACCG-----CAAGCCGAT 374
Db 238 ValArgGluArgLeuAlaLeuGlyArgArgGlyGlnArgArgGlnArgAlaGlu 257
QY 375 ---CAACCT-----CATGAGCTGGGGCGTCACCGACAACGTCATCAAGATGAGCCAGCT 425
Db 258 IleGlnProGlyThrHisArgProArgArgProArgHisAlaGlnGluAspArgProAla 277
QY 426 GAAGGACTTCAACTTCAA 443
Db 278 GlyAspLeuArgLeuArg 283

RESULT 11
US-10-425-114-56484
; Sequence 56484, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56484
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73213B10_FLI.pep
US-10-425-114-56484
```

```
Alignment Scores:
Pred. No.: 0.00352 Length: 347
Score: 119.00 Matches: 48
Percent Similarity: 42.47% Conservative: 14
Best Local Similarity: 32.88% Mismatches: 44
Query Match: 13.82% Indels: 40
DB: 4 Gaps: 9
```

US-10-049-750-13 (1-483) x US-10-425-114-56484 (1-347)

```
QY 75 GGAAGCCATGAAGCCCTCAAGAAAACCCCAAGTTCGATGACCTGGAACACAG----- 125
Db 178 GlyLeuHisGlyGlyArgArgAlaLeu-----ProGlyArgHisArgArgAla 193
QY 126 -----CTAGTTCCTCCCTGGACACCAAGGTATCCG 161
Db 194 AlaAlaArgGlnArgValProLeuLeuArgValProGlyGlnProAlaArgHisGlnPro 213
QY 162 GGTGTGATGAACACCCGGAATACCTGCATGACAAAGTTCGCTTACGCGCAC----- 212
Db 214 Gly-----LeuArgHisValProAla-----GlyHisHisArgGln 225
```

```
Qy 213 ----CTACAAACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGTGTCTACAT 269
Db 226 GlyGlnArgGlnArgProGlnLeuGlnGlnProValArgArgHisGlyGly-----242
Qy 270 CCCTGACGAAGACGCTGGCCCTGGCGATGGAACCTGGTTACGCTTGAGCCAAAGCAA 329
Db 243 -----ArgArgArgArgArgArgArgGluGlyArgArgGlnArgGlnArgArg 260
Qy 330 GTACGTCCTTTTGGTCTATCCCGGACGAGACTACGG-----CAAGCCGAT 374
Db 261 ValArgGluArgLeuAlaLeuGlyArgArgLeuArgGlyGlnArgArgGlnArgAlaGlu 280
Qy 375 ---CAACCT-----CATGAGCTGGGCGTCAGCGACACCTGATCAAGATGACCCAGCT 425
Db 281 IlegInProGlyThrHisArgProArgArgProArgHisAlaGlnGluAspArgProAla 300
Qy 426 GAAGGACTTCAACTTCAA 443
Db 301 GlyAspLeuArgLeuArg 306

RESULT 12
US-10-425-114-59602
; Sequence 59602, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59602
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3594-027-Al_FLI.pep
US-10-425-114-59602

Alignment Scores:
Pred. No.: 0.00356 Length: 361
Score: 119.00 Matches: 48
Percent Similarity: 42.47% Conservative: 14
Best Local Similarity: 32.88% Mismatches: 44
Query Match: 13.82% Indels: 40
DB: 4 Gaps: 9

US-10-049-750-13 (1-483) x US-10-425-114-59602 (1-361)
Qy 75 GGAAGCCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGGAAACAG-----125
Db 192 GlyLeuHisGlyGlyArgArgAlaLeu-----ProGlyArgHisArgAla 207
Qy 126 -----CTACGTTCCCTCGACACACGATACAGGTTATCCG 161
Db 208 AlaAlaArgGlnArgValProLeuLeuArgValProGlyGlnProAlaArgHisGlnPro 227
Qy 162 GGTTCATGACACCCGGAATACCTGCATGACAAAGTTTGGGCTACGGCCAC-----212
Db 228 Gly-----LeuArgHisValProAla-----GlyHisArgGln 239
Qy 213 ---CTACAAACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGTGTCTACAT 269
Db 240 GlyGlnArgGlnArgProGlnLeuGlnGlnProValArgArgHisGlyGly-----256
Qy 270 CCCTGACGAAGAACGCTGGCCCTGGGCATGGAACCTGACGCTTGACGCCAAGGCAA 329
```

```
Db 257 -----ArgArgArgArgArgAlaArgGluGlyArgArgAlaGlnArgGlnGlyArgArg 274
Qy 330 GTACGTCCTTTTGGTCTATCCCGGACGAGACTACGG-----CAAGCCGAT 374
Db 275 ValArgGluArgLeuAlaLeuGlyArgArgLeuArgGlyGlnArgArgGlnArgAlaGlu 294
Qy 375 ---CAACCT-----CATGAGCTGGGCGTCAGCGACACCTGATCAAGATGACCCAGCT 425
Db 295 IlegInProGlyThrHisArgProArgArgProArgHisAlaGlnGluAspArgProAla 314
Qy 426 GAAGGACTTCAACTTCAA 443
Db 315 GlyAspLeuArgLeuArg 320

RESULT 13
US-10-425-114-68177
; Sequence 68177, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68177
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3594-049-B7_FLI.pep
US-10-425-114-68177

Alignment Scores:
Pred. No.: 0.00356 Length: 362
Score: 119.00 Matches: 48
Percent Similarity: 42.47% Conservative: 14
Best Local Similarity: 32.88% Mismatches: 44
Query Match: 13.82% Indels: 40
DB: 4 Gaps: 9

US-10-049-750-13 (1-483) x US-10-425-114-68177 (1-362)
Qy 75 GGAAGCCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGGAAACAG-----125
Db 193 GlyLeuHisGlyGlyArgArgAlaLeu-----ProGlyArgHisArgAla 208
Qy 126 -----CTACGTTCCCTCGACACACGATACAGGTTATCCG 161
Db 209 AlaAlaArgGlnArgValProLeuLeuArgValProGlyGlnProAlaArgHisGlnPro 228
Qy 162 GGTTCATGACACCCGGAATACCTGCATGACAAAGTTTGGGCTACGGCCAC-----212
Db 229 Gly-----LeuArgHisValProAla-----GlyHisArgGln 240
Qy 213 ---CTACAAACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGTGTCTACAT 269
Db 241 GlyGlnArgGlnArgProGlnLeuGlnGlnProValArgArgHisGlyGly-----257
Qy 270 CCCTGACGAAGAACGCTGGCCCTGGGCATGGAACCTGACGCTTGACGCCAAGGCAA 329
Db 258 -----ArgArgArgArgArgAlaArgGluGlyArgArgAlaGlnArgGlnGlyArgArg 275
Qy 330 GTACGTCCTTTTGGTCTATCCCGGACGAGACTACGG-----CAAGCCGAT 374
Db 276 ValArgGluArgLeuAlaLeuGlyArgArgLeuArgGlyGlnArgArgGlnArgAlaGlu 295
```

```
QY 375 ---CAACCT-----CATGAGTCGGCGCTCAGCGCAACAGTGTATCAAGATGAGCCAGCT 425
Db 296 IleGlnProGlyThrHisArgProArgArgProArgHisAlaGlnGluAspArgProAla 315
QY 426 GAAGGACTCAACTTCAA 443
Db 316 GlyAspLeuArgLeuArg 321

RESULT 14
US-10-425-114-61578
; Sequence 61578, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 61578
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3069-035-G11_FLI.pep
US-10-425-114-61578

Alignment Scores:
Pred. No.: 0.00497 Length: 333
Score: 117.50 Matches: 48
Percent Similarity: 42.77% Conservative: 20
Best Local Similarity: 30.19% Mismatches: 58
Query Match: 13.65% Indels: 33
DB: 4 Gaps: 6

US-10-049-750-13 (1-483) x US-10-425-114-61578 (1-333)
QY 23 CGATCTACTTCGGTGGCGGTCTCACTGACCGCCCAAAACAAGCCT-----ACA 73
Db 107 ArgGlyThrSerThrGlySerArgThrCysAlaThrAlaProGlyCysGlyThr 126
QY 74 AGNAGCCATGGAAGCCCTCAAGGAAACCCACGATTGACCTGGAAACAGCTACGTTTC 133
Db 127 ArgAlaThrArgProSerArgSerSerThrArgSerProAlaArgArgSerThr 146
QY 134 CCCTGGACAACCAAGTACAGGGTATCGGGTTGATGAACACCCGGGATACCTGCATGACA 193
Db 147 ProArgGlyAlaThrProSerSerSerCysProThrThrArgThr----- 163
QY 194 AGGTTGGGTACGGCCACCTTACAACAACGACTTGAACGGGATCAAGCAACCAACGACATCA 253
Db 164 -----ArgThrSerProCysSerGlyProArgAlaAlaThrSer 176
QY 254 TGCTGGGTGTCTACATCCCTGACGAGAGACGTGCGCCTGGGGATCGAACTGGGTACG 313
Db 177 AlaThrAlaSerAlaAla-----SerAlaTrpSerThr----- 187
QY 314 CCTTGAGCCAAAGCAAGTACGTCCTTTTGGTTCATCCCGGACGAGACTACGCAAGC--- 370
Db 188 -----ThrSerThrSerThrSerThrProSerThrAlaThrArgLeuThrAlaAlaSer 205
QY 371 -----CGATCAACCTCATGAGCTGGGCGGTGACGCGACCAACGATGATCAAGATGA 418
Db 206 MetThrValProArgSerCysSerGlySerGlyAlaAlaArgAlaThr----- 221
QY 419 GCCAGCTGAGGACTTCAACTTCAACAAGCGCGCTTCGACTTCTACGAAAGTGCCG 475
```

```
Db 222 SerAlaGlyArgSerSerLeuGlyGluArgThrLysGluThrThrSerTrpTrpPro 240

RESULT 15
US-10-425-114-53657
; Sequence 53657, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53657
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-064-E8_FLI.pep
US-10-425-114-53657

Alignment Scores:
Pred. No.: 0.00518 Length: 383
Score: 117.50 Matches: 48
Percent Similarity: 42.77% Conservative: 20
Best Local Similarity: 30.19% Mismatches: 58
Query Match: 13.65% Indels: 33
DB: 4 Gaps: 6

US-10-049-750-13 (1-483) x US-10-425-114-53657 (1-383)
QY 23 CGATCTACTTCGGTGGCGGTCTCACTGACCGCCCAAAACAAGCCT-----ACA 73
Db 157 ArgGlyThrSerThrGlySerArgThrCysAlaThrAlaProGlyCysGlyThr 176
QY 74 AGNAGCCATGGAAGCCCTCAAGGAAACCCACGATTGACCTGGAAACAGCTACGTTTC 133
Db 177 ArgAlaThrArgProSerArgSerSerThrArgSerProAlaArgArgSerThr 196
QY 134 CCCTGGACAACCAAGTACAGGGTATCGGGTTGATGAACACCCGGGATACCTGCATGACA 193
Db 197 ProArgGlyAlaThrProSerSerSerCysProThrThrArgThr----- 213
QY 194 AGGTTGGGTACGGCCACCTTACAACAACGACTTGAACGGGATCAAGCAACCAACGACATCA 253
Db 214 -----ArgThrSerProCysSerGlyProArgAlaAlaThrSer 226
QY 254 TGCTGGGTGTCTACATCCCTGACGAGAGACGTGCGCCTGGGCATGGAAGTGGGTACG 313
Db 227 AlaThrAlaSerAlaAla-----SerAlaTrpSerThr----- 237
QY 314 CCTTGAGCCAAAGCAAGTACGTCCTTTTGGTTCATCCCGGACGAGACTACGCAAGC--- 370
Db 238 -----ThrSerThrSerThrSerThrProSerThrAlaThrArgLeuThrAlaAlaSer 255
QY 371 -----CGATCAACCTCATGAGCTGGGCGGTGACGCGACCAACGATGATCAAGATGA 418
Db 256 MetThrValProArgSerCysSerGlySerGlyAlaAlaArgAlaThr----- 271
QY 419 GCCAGCTGAGGACTTCAACTTCAACAAGCGCGCTTCGACTTCTACGAAAGTGCCG 475
Db 272 SerAlaGlyArgSerSerLeuGlyGluArgThrLysGluThrThrSerTrpTrpPro 290

Search completed: December 23, 2005, 23:20:16
Job time : 99.1442 secs
```

---

**THIS PAGE BLANK (USE)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:44:48 ; Search time 3.9562 Seconds  
(without alignments)  
1741.646 Million cell updates/sec

Title: US-10-049-750-13

Perfect score: 861

Sequence: 1 gtatacaaatgccaaaaa.....acgaagtgccgtatactaa 483

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 108002

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlpl  
-Q=/cgn2\_1/USPTO.spool.p/US10049750/runat.23122005.113555.6421/app.query.fasta.1.1742  
-DB=Published Applications AA New -QFMT=fastan -SUFFIX=n2p.rapbn -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10049750 @CGN 1.1.1 @runat.23122005.113555.6421  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New.\*

1: /cgn2\_6/ptodata/2/pubpaa/US08 NEW PUB.pcp.\*  
2: /cgn2\_6/ptodata/2/pubpaa/US06 NEW PUB.pcp.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US07 NEW PUB.pcp.\*  
4: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW PUB.pcp.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US09 NEW PUB.pcp.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US10 NEW PUB.pcp.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US11 NEW PUB.pcp.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US60 NEW PUB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	95	11.0	445	US-10-793-626-1294	Sequence 1294, Ap
2	94.5	11.0	1236	US-10-873-528-109	Sequence 109, App
3	92.5	10.7	5179	US-11-108-172-1068	Sequence 1068, Ap
4	87.5	10.2	1076	US-10-131-828A-219	Sequence 219, App
5	85.5	9.9	761	US-10-485-517-252	Sequence 252, App
6	79.5	9.2	596	US-11-102-240-100	Sequence 100, App
7	78	9.1	750	US-11-070-627-4	Sequence 4, Appli
8	76.5	8.9	957	US-11-108-172-1065	Sequence 1065, Ap
9	76	8.8	1532	US-10-821-234-914	Sequence 914, App
10	75.5	8.8	616	US-10-858-730-21	Sequence 21, Appli

11	75	8.7	537	6	US-10-641-678-62	Sequence 62, Appli
12	74.5	8.7	605	7	US-11-094-586-4	Sequence 4, Appli
13	74.5	8.7	672	7	US-11-000-463-455	Sequence 455, App
c 14	74.5	8.6	993	7	US-11-137-465-36	Sequence 36, Appli
15	74	8.6	431	6	US-10-821-234-1065	Sequence 1065, Ap
16	74	8.6	540	6	US-10-641-678-60	Sequence 60, Appli
17	73.5	8.5	168	7	US-11-044-111-27	Sequence 27, Appli
18	73.5	8.5	252	6	US-10-793-626-1948	Sequence 1948, Ap
19	73.5	8.5	276	6	US-10-467-657-4172	Sequence 4172, Ap
20	73.5	8.5	276	6	US-10-467-657-7078	Sequence 2078, App
21	73	8.5	756	7	US-11-074-176-202	Sequence 202, App
22	72.5	8.4	1464	7	US-11-000-463-243	Sequence 243, App
23	72.5	8.4	1464	7	US-11-186-284-28	Sequence 28, Appli
24	72.5	8.4	1467	6	US-10-821-234-1096	Sequence 1096, Ap
c 25	72.5	8.4	1798	6	US-10-995-561-1033	Sequence 1033, Ap
c 26	72.5	8.4	1798	6	US-10-995-561-1034	Sequence 1034, Ap
27	71.5	8.3	282	7	US-11-087-177-9	Sequence 9, Appli
28	71	8.2	305	7	US-11-080-091-2	Sequence 2, Appli
29	71	8.2	305	7	US-11-087-177-7	Sequence 7, Appli
30	71	8.2	390	6	US-10-485-517-235	Sequence 235, App
31	70.5	8.2	750	7	US-11-070-627-3	Sequence 3, Appli
32	70.5	8.2	1366	6	US-10-821-234-1431	Sequence 1431, Ap
33	70.5	8.2	1366	7	US-11-186-284-31	Sequence 31, Appli
c 34	70	8.1	52	6	US-10-467-657-4144	Sequence 4144, Ap
c 35	70	8.1	406	6	US-10-878-556A-41	Sequence 41, Appli
36	70	8.1	673	7	US-11-102-240-16	Sequence 16, Appli
c 37	70	8.1	744	7	US-11-186-284-37	Sequence 37, Appli
c 38	70	8.1	744	7	US-11-186-284-39	Sequence 39, Appli
39	70	8.1	1388	6	US-10-821-234-1143	Sequence 1143, Ap
40	69.5	8.1	391	6	US-10-979-821-12	Sequence 12, Appli
c 41	69.5	8.0	638	6	US-10-485-517-260	Sequence 260, App
42	69.5	8.1	750	7	US-11-070-627-2	Sequence 2, Appli
43	69.5	8.1	786	7	US-11-070-627-9	Sequence 9, Appli
c 44	69.5	8.0	1172	7	US-11-186-284-203	Sequence 203, App
c 45	69	8.0	277	7	US-11-132-285-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-10-793-626-1294

; Sequence 1294, Application US/10793626

; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1294

; LENGTH: 445

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: amino acid sequence

US-10-793-626-1294

Alignment Scores:

Pred. No.: 0.0472 Length: 445

Score: 95.00 Matches: 54

Percent Similarity: 43.82% Conservative: 24

Best Local Similarity: 30.34% Mismatches: 54

Query Match: 11.03% Indels: 46

DB: 6 Gaps: 10

US-10-049-750-13 (1-483) x US-10-793-626-1294 (1-445)

QY 5 ACTAATGCCAAAAGAGGATCTACTTCGGTCCGGCTGTCTACTGACCCCAAAACA 64





## RESULT 5

US-10-485-517-252  
; Sequence 252, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, Simon  
; APPLICANT: Mond, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629W0  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; PRIOR FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 252  
; LENGTH: 761  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-252

Alignment Scores:  
Pred. No.: 0.46 Length: 761  
Score: 85.50 Matches: 39  
Percent Similarity: 45.59% Conservative: 23  
Best Local Similarity: 28.68% Mismatches: 37  
Query Match: 9.93% Indels: 37  
DB: 6 Gaps: 5

US-10-049-750-13 (1-483) x US-10-485-517-252 (1-761)

```
QY 140 ACAACAGTACAGGGTATCGGGTTGATGAACACCCGGGAATACCTGCATGACAAGGTTT 199
   :::::::::::::::::::: :::: ::::
Db 344 SerThrSerThrSerLeuSerAsnSerAlaSerAlaSerGluSerAspLeuSerThr 363
   :::::::::::::::::::: :::: ::::
QY 200 GGGCTACGGCCACTCAACAAACGACTTGAACGGGATCAAGA-----CCAACGACA 250
   :::::::::::::::::::: :::: ::::
Db 364 SerLeuSerAspSerThrSerAlaSerMetGlnSerSerGluSerAspSerGlnSerThr 383
   :::::::::::::::::::: :::: ::::
QY 251 TCATGCTGGGTGTTACATCCCTGACGAAGAAGACGTCGGCCTGGGCATGGAACCTGGGTT 310
   :::::::::::::::::::: :::: ::::
Db 384 SerAlaSerLeuSerAspSerLeuSerThrSerThrSer-----AsnArgMet 399
   :::::::::::::::::::: :::: ::::
QY 311 ACGCCTTGACCAAGGCAAGTACGTCCTTTTGGTTCATCCCGGACGAAGACTACGGCA--- 367
   :::::::::::::::::::: :::: ::::
Db 400 SerThrIleAlaSerLeuSerThrSerValSerThrSerGluSerGlySerThrSerGlu 419
   :::::::::::::::::::: :::: ::::
QY 368 -----AGCCGA 373
   :::::::::::::::::::: :::: ::::
Db 420 SerThrSerGluSerAspSerThrSerThrSerLeuSerAspSerGlnSerThrSerArg 439
   :::::::::::::::::::: :::: ::::
QY 374 TCAACCTCATGAGCTGGGGGCTCAGCGACACACGTGATCAAGATGAGCCAGCTGAAGGACT 433
   :::::::::::::::::::: :::: ::::
Db 440 SerThrSerAlaSerGlySerAlaSerThrSer-----Thr 451
   :::::::::::::::::::: :::: ::::
QY 434 TCAACTTCAACAGCCCG-----GCTTCGACTTCTACGAAGGTG 472
   :::::::::::::::::::: :::: ::::
Db 452 SerThrSerAspSerArgSerThrSerAlaSerThrSerThrSerMet 467
   :::::::::::::::::::: :::: ::::
```

## RESULT 6

US-11-102-240-100  
; Sequence 100, Application US/11102240  
; Publication No. US20050260647A1  
; GENERAL INFORMATION:  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.

APPLICANT: Wood, William I.  
; TITLE OF INVENTION: ANTIBODIES TO POLYPEPTIDES ENCODED BY A NUCLEIC ACID UNDEREXPRESS  
; TITLE OF INVENTION: ESOPHAGEAL TUMOR  
; FILE REFERENCE: P3230R1C106C  
; CURRENT APPLICATION NUMBER: US/11/102,240  
; CURRENT FILING DATE: 2005-04-08  
; PRIOR APPLICATION NUMBER: 10/063662  
; PRIOR FILING DATE: 2002-05-07  
; PRIOR APPLICATION NUMBER: 10/006867  
; PRIOR FILING DATE: 2001-12-06  
; PRIOR APPLICATION NUMBER: PCT/US00/23328  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 60/170262  
; PRIOR FILING DATE: 1999-12-09  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-11-102-240-100

Alignment Scores:  
Pred. No.: 1.65 Length: 596  
Score: 79.50 Matches: 38  
Percent Similarity: 44.22% Conservative: 27  
Best Local Similarity: 25.85% Mismatches: 67  
Query Match: 9.23% Indels: 15  
DB: 7 Gaps: 4

US-10-049-750-13 (1-483) x US-11-102-240-100 (1-596)

```
QY 26 TCTACTTCGGTGCGGCTGTTTCACCTGACGCCCAAAACAAAGCCTACAAGGAAGCCATGG 85
   :::::::::::::::::::: :::: ::::
Db 170 SerSerThrLeuSerSerGlyAlaSerThrAlaThrAsnSerAspSer-----185
   :::::::::::::::::::: :::: ::::
QY 86 AAGCCCTCAAGGAAACCCCAACGATTGACCTCGGAAACACGACTAGCTTCCCTCGACAACC 145
   :::::::::::::::::::: :::: ::::
Db 186 -----SerThrThrSerSerGlyAlaSerThrAlaThrAsnSer---GluSer 200
   :::::::::::::::::::: :::: ::::
QY 146 AGTACAGGGTATCCGGGTTGATGAACACCCCGAATACCTGCATGACAAGGTTTGGGCTA 205
   :::::::::::::::::::: :::: ::::
Db 201 SerThrThrSerSerGlyAlaSerThrAlaThrAsnSerGluSerSerThrValSerSer 220
   :::::::::::::::::::: :::: ::::
QY 206 CGCCACCTACAACACGACTTGAACGGGATCAAGCAACCAACGACATCATCTGGGTGCTCT 265
   :::::::::::::::::::: :::: ::::
Db 221 ArgAlaSerThrAlaThrAsnSerGluSerSerThrThrSerSerGlyAlaSerThrAla 240
   :::::::::::::::::::: :::: ::::
QY 266 ACATCCCTGACGAAGAGAGCTCGGCTGGGCATGGNACTGGGTAGCCCTTGAGCCCAAG 325
   :::::::::::::::::::: :::: ::::
Db 241 ThrAsnSerGluSerArgThrThrSerAsnGlyAlaGlyThrAlaThrAsn---SerGlu 259
   :::::::::::::::::::: :::: ::::
QY 326 GCAAGTACGTCCTTTTGGTTCATCCCGGACCAAGACTACGCAAGCCGATCAACCTCATGA 385
   :::::::::::::::::::: :::: ::::
Db 260 SerSerThrThrSerSerGlyAlaSerThrAlaThrAsnSerAspSerSerThrValSer 279
   :::::::::::::::::::: :::: ::::
QY 386 GCTGGGGCGTCAGCGACACGATCAAGATGAGCCAGCTGAAGGACTTCAACTTCAACA 445
   :::::::::::::::::::: :::: ::::
Db 280 SerGlyAlaSerThrAlaThrAsnSer-----GluSerSerThrThrSer 294
   :::::::::::::::::::: :::: ::::
QY 446 AGCGGCGCTTCGACTTCTACG 466
   :::::::::::::::::::: :::: ::::
Db 295 SerGlyAlaSerThrAlaThr 301
   :::::::::::::::::::: :::: ::::
```

## RESULT 7

US-11-070-627-4  
; Sequence 4, Application US/11070627  
; Publication No. US20050271625A1  
; GENERAL INFORMATION:  
; APPLICANT: Nash, Kevin R.  
; APPLICANT: Burger, Corinna  
; TITLE OF INVENTION: RAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE  
; FILE REFERENCE: 36689.8  
; CURRENT APPLICATION NUMBER: US/11/070,627



```
QY 326 GCAAGTAGCTCCTTTGGTTCATCCCGACGACGACTACGGCA-----AGCCGATCA 376
Db |||||
723 AlaserThrGlnThrGlyLeuProAlaThrLeuThrThrAlaAspLeuGlyGluGluser 742
QY 377 ACCTCATGAGCTGGGGCGTCAGCGACCAACGTCATCAAGATGAGCGACCTGAAGGACTTCA 436
Db |||||
743 ThrThrPheProSerSerSerGlySerThr-----GlyThrThr 755
QY 437 ACTTCAACAGCCGCTTCGACTTCT 463
Db |||||
756 LeuSerProAlaArgSerThrThrSer 764
RESULT 9
US-10-821-234-914
; Sequence 914, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR FILING DATE: 2004-04-07
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 914
; LENGTH: 1532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-914
Alignment Scores:
Pred. No.: 4.7 Length: 1532
Score: 76.00 Matches: 35
Percent Similarity: 42.11% Conservative: 13
Best Local Similarity: 30.70% Mismatches: 44
Query Match: 8.83% Indels: 22
DB: 6 Gaps: 7
US-10-049-750-13 (1-483) x US-10-821-234-914 (1-1532)
QY 60 AAACAAGCCTCAACAAGAGCCATCGAAGCCCTCAAGAAAAACCAACGATTGACCTGGA 119
Db |||||
590 LysLeuMetMetGluGlnGluAenGlyAsnLeuArgGlySerProGly-----ProLys 607
QY 120 AACAGCTACGTTCCCTCTGGA---CAACCAAGTCAAGGGTATCGGGTTGATGAACACCC 176
Db |||||
608 GlyAspMetGlySerProGlyProLysGlyAspArgGlyPheProGly-----ThrPro 625
QY 177 GGAATACCTGCATGACAAAGTTTGGGCTAC-----GCCACCTACACACGA 224
Db |||||
626 GlyIleProGly-----ProLeuGlyHisProGlyProGlnGlyProLysGlyGlnLys 643
QY 225 -----CTTGAACGGGATCAAGACCAACGACATCATGCTGGGTGT 263
Db |||||
644 GlySerValGlyAspProGlyMetGluGlyProMetGlyGlnArgGlyArgGluGlyPro 663
QY 264 CTACATCCCTGAGAAAGACGTCGGCCTGGGATGGA-----ACTGGGTGA 311
Db |||||
664 MetGlyProArgGlyGluAlaGlyProProGlySerGlyGluLysGlyGluArgGlyAla 683
QY 312 CGCCTTGAGCCAAAGGCAAGTACGTCCTTTTGGTCATCCCGGA 353
Db |||||
684 AlaGlyGluProGlyProHisGlyProProGlyValProGly 697
RESULT 10
US-10-858-730-21
; Sequence 21, Application US/10858730
; Publication No. US20050255568A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 616
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-10-858-730-21
Alignment Scores:
Pred. No.: 4.09 Length: 616
Score: 75.50 Matches: 40
Percent Similarity: 34.72% Conservative: 10
Best Local Similarity: 27.78% Mismatches: 42
Query Match: 8.77% Indels: 52
DB: 6 Gaps: 7
US-10-049-750-13 (1-483) x US-10-858-730-21 (1-616)
QY 14 CAAAAAGACGATCTACTTCGTCGCGCTTCTCACTGACCCCAAAACAAAGCCTTACA 73
Db |||||
71 GlnArgArgAlaAspAlaValProProGlyArgGlnAlaLeuArgGluArgHis 90
QY 74 AGG-----AAGCCATGGAAGCCCTCAAGAAAAACCAACGATTGACCTGGA----- 119
Db |||||
91 ArgAlaArgProLeuProCysArgProAlaSerArgArg-ProGlySerSerG1 110
QY 120 -----AAACAGTACGTTCCCTCGACCAACCAAGGTATCCGGTTTCATG 169
Db |||||
110 YArgHisArgArgLeuLeuHisGlyGlnGlnLeuGln----- 124
QY 170 AACACCCGGAATACCTGCATGACAAGGTTTGGGCTACGGCCACCTACAACAACGACTTGA 229
Db |||||
125 ----ProArgAlaProAlaCysArgGlyArgGly----- 134
QY 230 ACGGGATCAAGACCAACGACATCATGCTGGGTGTCTACATCCCTGACGAAGACGCTCG 289
Db |||||
135 -----ProArgGluGluArgProAr 141
QY 290 GCCTGGGCATGGAACCTGGGTTACGCCTTGACCAAGCAAGTACGTCCTTTTGGTCATCC 349
Db |||||
141 gProGly---AlaThrGlyAsnArgArgArgProValAlaAla----- 154
QY 350 CGGACCAAGATACCGCAAGCCGATCAACCTCATGAGCTGGGGCGTCAGCGACAACGTGA 409
Db |||||
155 -GlyArgArgLeuSerSer-----GlyArgArgSerGlyHi 167
QY 410 TCAAGATGAG 419
Db |||||
167 sHisAspGlu 170
RESULT 11
US-10-641-678-62
; Sequence 62, Application US/10641678
```







```
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pc_seq_genes Version 1.0
; SEQ ID NO 1065
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1065

Alignment Scores:
Pred. NO.: 5.19 Length: 431
Score: 74.00 Matches: 42
Percent Similarity: 42.04% Conservative: 24
Best Local Similarity: 26.75% Mismatches: 63
Query Match: 8.59% Indels: 28
DB: Gaps: 8

US-10-049-750-13 (1-483) x US-10-821-234-1065 (1-431)
Qy 35 GTGGCGGTGTTCACTACGCCCAACAAAGCCTACAGGAAGCCATGGAAGCCCTCA 94
Db ::::||||| :::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 72 LeuProSerGlyPheMetSerLeuAspAsnGlyThrAlaThrProGluLeuProThr 91
Db ::::||||| :::: ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 95 AGGAAACCC-----CAACGATTGACCTGGAAACACAGCTACGTTCCCTGG 139
Db :::: ||| :::: ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 92 GlnGlyThrPheSerAsnValSerThrAsnValSerTyGlnGluThrThrProSer 111
Db :::: ||| :::: ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 140 ACAACC---AGTACAAAGGTATCCGGTTGTATGAACACCCGGAATACCTGCATGACAAGG 196
Db :||| ||| ||| ||| :::: ||| ||| ||| ||| ||| ||| |||
Qy 112 ThrLeuGlySerThrSerLeuHisProValSerGlnHisGlyAsnGluAlaThrThrAsn 131
Db :||| ||| ||| ||| :::: ||| ||| ||| ||| ||| ||| |||
Qy 197 TTTGGGCTACGGCCACCTACACACGACTTCAACGGGATCAAGACCAACGACATATGC 256
Db :||| ||| ||| ||| :||| ||| ||| ||| ||| ||| ||| |||
Qy 132 Ile-----ThrGluThrThrVallyspheThrSerThrSerValIleThrSerVal 148
Db :||| ||| ||| ||| :||| ||| ||| ||| ||| ||| ||| |||
Qy 257 TGGGTGTCTACA-----TCCCTGACGAAGAGAGCTCGGCTGGGCATGGAAGTGGGTT 310
Db :::: ||| ||| ||| :||| ||| ||| ||| ||| ||| ||| |||
Qy 149 TyrGlyAsnThrAsnSerSerValGlnSerGlnThrSerValIleSerThrValPheThr 168
Db :||| ||| ||| ||| :||| ||| ||| ||| ||| ||| ||| |||
Qy 311 ACGCTTGAGCCCAAGCAAGTACGTCTTTTGGTTCATCCCGACGACGACTACGGCAAGC 370
Db :||| ||| ||| ||| :||| ||| ||| ||| ||| ||| ||| |||
Qy 169 ThrPro---AlaAsnValSerThr-----ProGluThrThrLeuLys 181
Db :||| ||| ||| ||| :||| ||| ||| ||| ||| ||| ||| |||
Qy 371 CGATCAACTCATGAGCTGGGCGCTCAGGCACACGTCATCAAGATGAGCCAGCTGAAGG 430
Db :||| ||| ||| ||| :||| ||| ||| ||| ||| ||| ||| |||
Qy 182 ProSerLeuSerProGlyAsnValSerAspLeu----- 192
Db :||| ||| ||| ||| :||| ||| ||| ||| ||| ||| ||| |||
Qy 431 ACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGGTCCCGTATACT 481
Db ::::||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 193 SerThrThrSerThrSerLeuAlaThrSerProThrLys---ProTyrThr 208
Db ::::||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: December 23, 2005, 23:20:41
Job time : 12.9562 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:20:36 ; Search time 20.985 Seconds  
(without alignments)  
4429.126 Million cell updates/sec

Title: US-10-049-750-13  
Perfect score: 861  
Sequence: 1 gatactaataatgccaaaaa.....acgaaggtgcgtatactaa 483

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlp  
-O=/cgn2 1/uspto spool p/us10049750/runat 23122005 113553 6303/app query.fasta\_1.1742  
-DB=PIR -OPMT=fastan -SUFFIX=n2p.rpx -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORES=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10049750 @cgn 1 1.83 @runat 23122005 113553 6303 -NCPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	637	74.0	158	JC7522	nucleoside deoxyri
2	150.5	17.5	263	S01360	salivary glue prot
3	142	16.5	1832	T31113	mucin-like glycop
4	136.5	15.9	159	E86686	hypothetical prote
5	134.5	15.6	1161	S57180	probable membrane
6	133.5	15.5	354	T46740	microfilarial shea
7	128	14.9	477	S53362	mucin SAC (clone J
8	125.5	14.6	605	S248940	hypothetical prote
9	124	14.4	510	H84824	En/Spm-like transp
10	121.5	14.1	866	T45462	membrane glycoprot
11	121	14.1	662	A45155	mucin FIM-C.1 - Af
12	121	14.1	797	VG8EX1	glycoprotein X pre
13	116.5	13.5	851	T22696	hypothetical prote
14	116	13.5	232	A60095	larval glue protei

15	114	13.2	867	2	T45463	membrane glycoprot
16	113.5	13.2	770	2	T22808	hypothetical prote
17	111	12.9	279	2	S53363	mucin SAC (clone J
18	110.5	12.8	496	2	E90181	hypothetical prote
19	109.5	12.7	750	2	T42614	probable envelope
20	109	12.7	725	2	A41258	a-agglutinin core
21	108.5	12.6	796	2	T21460	hypothetical prote
22	108	12.5	371	2	S20075	promastigote surfa
23	106.5	12.4	402	2	E86185	hypothetical prote
24	105.5	12.3	327	2	S20074	promastigote surfa
25	105	12.2	328	2	S67570	hypothetical prote
26	105	12.2	815	2	JN0689	glutenin, high-mol
27	104.5	12.1	503	2	S63257	probable membrane
28	104.5	12.1	848	2	S02262	glutenin high mole
29	102.5	11.9	164	2	I53641	mucin SAC - human
30	102.5	11.9	838	1	EEWTHW	glutenin, high mol
31	102	11.8	507	2	T44768	antifreeze glycop
32	102	11.8	660	2	JW0067	chitinase (EC 3.2.
33	102	11.8	815	2	B30843	glutenin high mole
34	102	11.8	1367	1	S48478	glucan 1,4-alpha-g
35	101	11.7	328	2	S01359	salivary glue prot
36	101	11.7	1360	2	T33922	hypothetical prote
37	100.5	11.7	417	2	T24618	hypothetical prote
38	100	11.6	1777	2	T34369	hypothetical prote
39	99.5	11.6	339	2	T25562	hypothetical prote
40	99.5	11.6	379	2	S50125	larval glue protei
41	99.5	11.6	705	2	S18733	glutenin high mole
42	98.5	11.4	217	2	S01358	salivary glue prot
43	98.5	11.4	708	2	T19474	hypothetical prote
44	98.5	11.4	790	2	T34293	hypothetical prote
45	98.5	11.4	798	2	T34248	hypothetical prote

ALIGNMENTS

RESULT 1

JC7522  
nucleoside deoxyribosyltransferase (EC 2.4.2.6) II - Lactobacillus helveticus  
C:Species: Lactobacillus helveticus  
C>Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 09-Jul-2004  
C:Accession: JC7522; PC7103  
B:Okuyama, K.; Noguchi, T.  
Biosci. Biotechnol. Biochem. 64, 2243-2245, 2000  
A:Title: Molecular cloning and expression of the nucleoside deoxyribosyltransferase-II 9  
A:Reference number: JC7522; MUID:21012342; PMID:11129605  
A:Accession: JC7522  
A:Molecule type: DNA  
A:Residues: 1-158 <OKU>  
A:Cross-references: UNIPROT:Q9KWF0; UNIPARC:UPI0000015C9E3; DDBJ:AB039914  
A:Experimental source: strain ATCC 8018  
A:Accession: PC7103  
A:Molecule type: protein  
A:Residues: 1-11;53-62 <OK2>  
A:Cross-references: UNIPARC:UPI0000017A510; UNIPARC:UPI0000017A511  
C:Comment: This enzyme catalyzes transfer of glycosyl residues from a donor deoxynucleos

C:Keywords: glycosyltransferase; pentosyltransferase

Alignment Scores:	5.64e-50	Length:	158
Pred. No.:	637.00	Matches:	114
Score:	87.10%	Conservative:	21
Percent Similarity:	73.55%	Mismatches:	20
Best Local Similarity:	73.98%	Indels:	0
Query Match:	2	Gaps:	0
DB:			

US-10-049-750-13 (1-483) x JC7522 (1-158)

Qy 16 AAAAAGACGATCTACTTCGGTCCCGCTGGTTCACCTGACCGCAAAACAAAGCCTACAAG 75  
Db 4 LysLysThrLeuTyPheGlyAlaGlyTrpPheAsnGluLysGlnAsnLysalaTyLys 23







A:Map position: 8L  
C;Keywords: transmembrane protein

## Alignment Scores:

Pred. No.: 0.00242 Length: 605  
Score: 125.50 Matches: 51  
Percent Similarity: 45.96% Conservative: 23  
Best Local Similarity: 31.68% Mismatches: 48  
Query Match: 14.58% Indels: 39  
DB: 2 Gaps: 7

US-10-049-750-13 (1-483) x S48940 (1-605)

Qy 5 ACTAAATGCCAAAAGACGATCTACTTCGGTGGCGGTGGTTCACCTGACCGCCAAACA 64  
Db 181 ThrThrThrSerThrLysLeuSerThrSerLeuProThr---SerThrThrSerThr 199  
Qy 65 AAGCCTACAAGGAGCCATGGAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAACA 124  
Db 200 SerThrThrThrSerThrSerThrSerThrThrValSerValThrSerThr 219  
Qy 125 GCTACGTTCCCTGGACACGATACAGAGGTATCCGGGTGTGATGAACACCCGGAATACC 184  
Db 220 SerThr-----ThrThrSerThrThrSerThrLeuLeSerThr----- 233  
Qy 185 TGCATGACAGGTTGGCTACGGCCACCTACACACGACTTGAACGGGATCAAGACCA 244  
Db 234 -----SerThrSerSerSerSerThrPro 243  
Qy 245 ACGACA-----TCATGCTGGGTGTCTATCCCTGACGAGAAAGACGCTCGGCTGGGCA 298  
Db 244 ThrThrThrSerSerAlaProIleSerThrSerThrSerThrSerThr 261  
Qy 299 TGGAACTGGGTACGCCCTTGAGCCAAAGCAAGTACGTCCTTTTGGTCAATCCCGACGAAG 358  
Db 262 -----SerThrSerThrSerProThrSerSerSerAlaPro 274  
Qy 359 ACTAGCGCAACCGGATCAACCTCATGACCTGGGGCTGAGGACACGATCATCAAGATGA 418  
Db 275 ThrSerSerSerAsnThr-----ProThrSerThrThr 286  
Qy 419 GCCAGCTGAAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGTGGCGTAT 478  
Db 287 -----PheThrThrThrSerProSerThrAlaProSerSerThrValThrTyr 303  
Qy 479 ACT 481  
Db 304 Thr 304

## RESULT 9

H84824  
En/Spm-like transposon protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: H84824  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; PMID:20083487; PMID:10617197  
A;Accession: H84824  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-510 <STO>  
A;Cross-references: UNIPROT:004210; UNIPARC:UPI00000A9658; GB:AE002093; NID:g2088658; PI  
C;Genetics:  
A;Map position: 2

Alignment Scores:  
Pred. No.: 0.00327 Length: 510  
Score: 124.00 Matches: 51

Percent Similarity: 48.68% Conservative: 23  
Best Local Similarity: 33.55% Mismatches: 56  
Query Match: 14.40% Indels: 22  
DB: 2 Gaps: 6

US-10-049-750-13 (1-483) x H84824 (1-510)

Qy 32 TCGGTGCGCGTGGT-----TCACCTGACCGCCAAAACAAAGCCCTACAAGGAAGCCA 82  
Db 160 AlaThrProThrGlyArgSerSerThrLeuThrAlaAsnSerLysSerSerArg----- 177  
Qy 83 TGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAAAACAGCTACGTTCCCTCGACA 142  
Db 178 -----ProSerThrProThrSerArgAlaThrValSerSerAlaThrArgProSerLeu 195  
Qy 143 ACAGTACAGAGGTATCCGGGTGTGATGAACACCCGGGAATACCTGCATGACAAAGGTTGGG 202  
Db 196 ThrAsnSerArgSerThrValSerAlaThrThrLysProThrProMetSerArgSerThr 215  
Qy 203 CTACGGCCACCTACAACACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGGTG 262  
Db 216 SerLeuSerSerSerArgLeuThrProThrAlaSerLysProThrThrSerThrAlaArg 235  
Qy 263 TCTACA---TCCCTGACGAGAAAGACGCTCGGCTGGGCATGGAACCTGGGTTACGCTTGA 319  
Db 236 SerAlaGlySerValThrArgSerThrPro----- 245  
Qy 320 GCCAAGCAAGTACGTCCTTTTGGTCAATCCGGACGAGACTACGGCAAGCCGATCAACC 379  
Db 246 SerThrThrThrLysSerAlaGlyProSerArgSerThrThrProLeuSerArgSerThr 265  
Qy 380 TCATGACTGGGGCTGAGGACACGCTGATCAAGATGAGCAGCTCAAGGACTTCAACT 439  
Db 266 Ala-----ArgSerSerThrProThrSerArg---ProThrLeuProProSerLys 281  
Qy 440 TCAACAAAGCGCGTTCGACTTCTACGAAGGTGGCG 475  
Db 282 ThrIleSerArgSerSerThrProThrArgArgPro 293

## RESULT 10

T45462  
membrane glycoprotein [imported] - equine herpesvirus 1  
C;Species: equine herpesvirus 1  
C;Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T45462  
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.  
J. Equine Sci. 7, 79-87, 1996  
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equin  
A;Reference number: 222973  
A;Accession: T45462  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-866 <KIR>  
A;Cross-references: UNIPROT:O39781; UNIPARC:UPI00000ECBA1; EMBL:D88733; PIDN:BAA20037.1  
A;Experimental source: strain HHL  
C;Genetics:  
A;Note: ORF71  
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homo  
F;558-866/Domain: equine herpesvirus 1 glycoprotein homology <EHG>

Alignment Scores:  
Pred. No.: 0.00578 Length: 866  
Score: 121.50 Matches: 43  
Percent Similarity: 41.72% Conservative: 25  
Best Local Similarity: 26.38% Mismatches: 84  
Query Match: 14.11% Indels: 11  
DB: 2 Gaps: 2

US-10-049-750-13 (1-483) x T45462 (1-866)

Qy 26 TCTACTCGGTGCGGCTGGTTCACGT-----ACCGCAAAACAAAGCCCTACAAGGAAG 79  
Db 97 ThrThrSerIleProThrSerThrSerThrGluThrThrThrProThrAlaSer 116

```

QY      80 CCATGGAGCCCTCAGGAAACCCCAACGATTGACCTGGAAACAGCTAGCTTCCCTGG 139
Db      117 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 136
QY      140 ACAACAGTACAAGGTTATCCGGTTGATGATCAACACCCGGGAATACCTGCATGACAAGTTT 199
Db      137 ThrThrAlaAlaSerThrAlaAlaSerThrSerAlaGluThrThrThrThrThrThrThr 156
QY      200 GGGCTACGGCCACCTCAACAACAGCATTGAACGGGATCAAGACCAACGACA----- 250
Db      157 AlaThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 176
QY      251 -----TCACTGGGTCTCTACATCCCTGAGCAAGAGAGAGCTGGCC 292
Db      177 ThrValProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 196
QY      293 TGGGCATGGAACCTGGGTTAGCCTTGAGCCCAAGCAAGTACGCTTTTGGTCATCCGG 352
Db      197 AlaThrThrThrAlaAlaThrThrThrThrAlaAlaThrThrThrThrThrThrAla 216
QY      353 ACGAAGACTACGGCAAGCCGATCAACCTCATGAGCTGGGCGTCAGCGACAAGCTGATCA 412
Db      217 AlaThrThrThrAlaAlaThrThrThrThrAlaAlaThrThrThrThrThrThrAla 236
QY      413 AGATGAGCCAGCTGAAGGACTCAACTTCAACAGCGCGCTTCGACTTCTACGAAGGTG 472
Db      237 AlaThrThrThrAlaAlaThrThrSerSerAlaThrThrAlaAlaThrThrThrAlaAla 256
QY      473 CCGTATACT 481
Db      257 ThrThrThr 259

RESULT 11
A45155
mucin FIM-C.1 - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C>Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Jul-2004
C:Accession: A45155
R:Hauser, F.; Hoffmann, W.
J. Biol. Chem. 267, 24620-24624, 1992
A:Title: P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1)
A:Reference number: A45155; MUID:93077556; PMID:1447205
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-662 <HAU>
A:Cross-references: UNIPROT:Q05049; UNIPARC:UPI000012F8CF; GB:L02115; NID:g214147; PIDN:
F;162-202/Domain: trefoil homology <TRF1>
F;307-347/Domain: trefoil homology <TRF2>
F;354-394/Domain: trefoil homology <TRF3>
F;526-566/Domain: trefoil homology <TRF4>
F;573-613/Domain: trefoil homology <TRF5>
F;621-661/Domain: trefoil homology <TRF6>

Alignment Scores:
Pred. No.: 0.00627 Length: 662
Score: 121.00 Matches: 44
Percent Similarity: 47.52% Conservative: 23
Best Local Similarity: 31.21% Mismatches: 58
Query Match: 14.05% Indels: 16
DB: 2 Gaps: 4

US-10-049-750-13 (1-483) x A45155 (1-662)

QY      47 TCACGTACCCGCAAAACAAAGCCCTACAGGAGCGCTGAGCCCTCAGGAACCCAA 106
Db      396 SerGlnValAlaAlaThrLysThrThrThrThrThrThrThrThrThrThrThrThrThr 415
QY      107 CGATTGACCTGGAAA-----ACAGCTACGTTCCCTCGGCAACCAAGTACAAAGGGTATCCGGG 163
Db      416 ThrThrThrThrLysAlaAlaThrThrThrThrThrThrThrThrThrThrThrThrThr 431

```

```

QY      164 TTGATGAACACCGGAATACCTGCATGACAGGTTTGGGCTACGGCCACCTACAACAACG 223
Db      432 -----ThrProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 448
QY      224 ACTTGAAACCGGATCAAGACCAACGACATCATGCTGGGTGCTACATCCCTCGACGAAGAAG 283
Db      449 ThrProThrThrThrThrThrThrThrThrThrThrThrLysAlaAlaThrThrThrProThr 468
QY      284 AGTCGGCCTGGGCATGGAACCTGGGTTACGCTTGAGCCCAAGCAAGTACGTCCTTTTGG 343
Db      469 ThrThrThr-----ThrThrProThrThrThrThrThrThrLysAlaAlaThrThr 483
QY      344 TCATCCCGACGACGACTAGCGCAAGCCGATCAACCTCATGAGCTGGGCGTCAGCGACA 403
Db      484 ThrThrProThrThrThrThrThrThrThrProThrThrThrThrThrThrLysAlaAlaThrThr 503
QY      404 ACCTGATCAAGATGAGCCAGCTGAAGACTTCAACTTCAACAGCGCGCTTCGACTTCT 463
Db      504 ThrProThr-----ThrThrThrThrThrThrThrThrThrThrThrLysAlaAlaThrThr 520
QY      464 ACG 466
Db      521 Thr 521

RESULT 12
VGBEX1
glycoprotein X precursor - equine herpesvirus 1 (strain Ab4p)
C:Species: equine herpesvirus 1
A:Note: host Equus caballus (domestic horse)
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: H36802
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
submitted to GenBank, March 1992
A:Description: The DNA sequence of equine herpesvirus-1.
A:Reference number: A36805
A:Accession: H36802
A:Molecule type: DNA
A:Residues: 1-797 <TEL>
A:Cross-references: UNIPROT:P28968; UNIPARC:UPI0000138750; GB:M86664; NID:g330791; PIDN:
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Virology 189, 304-316, 1992
A:Title: The DNA sequence of equine herpesvirus-1.
A:Reference number: A41831; MUID:92295586; PMID:1318606
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Genetics:
A:Gene: 71
C:Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homolog
C:Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-797/Product: glycoprotein X #status predicted <MAT>
F;23-465/Region: serine/threonine-rich
F;489-797/Domain: equine herpesvirus 1 glycoprotein homology <EHG>
F;766-790/Domain: transmembrane #status predicted <TMN>
F;590/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 0.00637 Length: 797
Score: 121.00 Matches: 44
Percent Similarity: 44.65% Conservative: 27
Best Local Similarity: 27.67% Mismatches: 76
Query Match: 14.05% Indels: 12
DB: 1 Gaps: 3

US-10-049-750-13 (1-483) x VGBEX1 (1-797)

QY      26 TCACCTTCGTGGCGGTGTTCACTGACCGCAAAACAAGCCTACAAGAAGCCATGG 85
Db      97 ThrThrSerIleProThrSerThrSerThrGluThrThrThrThrThrThr----- 112
QY      86 AAGCCTCAAGGAAACCCCAACGATTGACCTGGAACACAGCT-----ACGTTCCCC 136
Db      113 ---ProThrAlaSerThrThrThrThrThrThrThrAlaAlaProThrThrAlaAla 131

```





```
Db 94 ThrThrThrThrProCysProThrThrThr-----ThrThrArgThr 107
Qy 182 ACCTGCATGCAAGGTTTGGGTACGGCCACTACAAACGACGATTGAACGGGATCAAGA 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 Thr-----ThrThrProThrThrThrThrArgThrThrThrThr 120
Qy 242 CCAACGACATCATGCTGGGTGTCTACATCCTCGACGAGAGAGCTGGCGCTGGCATGG 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ProThrThrThrThrArgThrThrThrThrProThrThrThrThr 135
Qy 302 AACTGGGTACGCGTTGAGCCAAAGCAAGTACGTCCTTTTGGTCATCCCGACGACGACT 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 -----ArgThrThrThrThrArgThrThrThrThrProThr 147
Qy 362 ACGCAAGCCGATCAACCTCATGAGCTGGGGCGTCAAGCAGACGTCGATCAAGATGAGCC 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 ThrThrThrArgThrThrThr-----ThrProThrThr 158
Qy 422 AGCTGAAGGACTTCAACTTCAACAGCGCGCTTCGACTTCTAGGAAGGTGCGGTACT 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 -----ThrThrProThrThrThrThrGlnThrThrThrArgAlaProProThr 175

RESULT 15
T45463
membrane glycoprotein [imported] - equine herpesvirus 1
C;Species: equine herpesvirus 1
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C;Accession: T45463
R;Kirisawa, R.; Kobayashi, T.; Kawakami, Y.; Iwai, H.
J. Equine Sci. 7, 79-87, 1996
A;Title: Nucleotide sequences of open reading frames 1, 24 and 71 of an attenuated equine
A;Reference number: Z22973
A;Accession: T45463
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-867 <KIR>
A;Cross-references: UNIPROT:O39782; UNIPARC:UPT00000EF109; EMBL:D88734; PIDN:BAA20038.1
A;Experimental source: isolate 3F clone; strain BK343
C;Genetics:
A;Note: ORF71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homodimer
```

```
Alignment Scores:
Pred. No.: 0.0279 Length: 867
Score: 114.00 Matches: 43
Percent Similarity: 44.03% Conservative: 27
Best Local Similarity: 27.04% Mismatches: 73
Query Match: 13.24% Indels: 16
DB: 2 Gaps: 4
```

US-10-049-750-13 (1-483) x T45463 (1-867)

```
Qy 26 TCTACTTCGGTGGCGGTGGTTCACGCGCCAAACAAAGCCTACAAGGAGCCATGG 85
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 ThrThrSerIleProThrSerThrSerThrGluThrThrThrThr----- 112
Qy 86 AAGCCCTCAAGGAAACCAACGATTCACCTGGGAAACAGCT-----ACGTTCCCC 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 ---ProThrAlaSerThrThrThrProThrThrThrAlaAlaProThrThrAlaAla 131
Qy 137 TGGACAACAGATACAGGGTATCCGGGTGATGATGAACACCCGGGAATACCTGCATGCAAGG 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 ThrThrThrAlaValThrThrAlaAlaSerThrSerAlaGluThrThrThrAlaThrAla 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 197 TTGGGGTACGGCCACCTACACACGACTTGACGGGATCGACCAAGCATCATGC 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 ThrAlaThrSerThrProThrThrThrProThrSerThrThrThrThrThrAlaThr 171
Qy 257 -----TGGGTGCTTACATCCCTGACGAGAGAGACGTCGGCGCTGGGCATGGAAC 304
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 ThrThrValProThrThrAlaSerThrThrThrThrThrThrThrThrThrThrThr 191
Qy 305 TGGGTTAGCGCTGAGCAAGCAAGTACGTCTTTTGGTTCATCCCGGACGAGACTACG 364
```

```
Db 192 AlaAlaThrThrThrAlaAlaThrThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 365 GCAAGCCGATCAACCTCATGAGCTGGGGCGTCAAGCAGCAACGTCGATCAAGATGAGCCAGC 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 AlaAlaThrThrThrAlaAlaThrThrThrThrAlaAlaThrThr-----ThrAla 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 425 TGAAGGACTTCAACTTCAACAGCCGCGCTTCGACTTCTACGAAGGTGCGGTATACT 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 228 AlaThrThrThrAlaAlaThrThrThrThrAlaAlaThrThrThrAlaAlaThrThrThr 246

Search completed: December 23, 2005, 22:46:45
Job time : 27.985 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: December 23, 2005, 22:15:36 ; Search time 130.382 Seconds  
(without alignments)  
5227.239 Million cell updates/sec

Title: US-10-049-750-13

Perfect score: 861

Sequence: 1 gtatacaaatgccaaaaa.....acgaaggtgcgtatactaa 483

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool\_p/US10049750/runat\_23122005.113552.6292/app\_query.fasta\_1.1742  
-DB=UniProt -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10049750@cgn 1.1 614 @runat\_23122005.113552.6292 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	843	97.9	156	1 NTD_LACLE	Q9r5v5 lactobacill
2	728	84.6	158	2 Q8RLY4 LACHE	Q8rly4 lactobacill
3	719	83.5	159	2 Q5FMH8 LACAC	Q5fmm8 lactobacill
4	653	75.8	149	1 NTD_LAGJO	Q7alg9 lactobacill
5	634	73.6	158	2 Q9KWF0 LACHE	Q9krf0 lactobacill
6	209	24.3	168	1 NTD_LACFE	Q6yni5 lactobacill
7	171.5	19.9	167	2 Q8RLY5 LACHE	Q8rly5 lactobacill
8	154.5	17.9	167	2 Q5FIN0 LACAC	Q5fin0 lactobacill
9	150.5	17.5	263	1 SGS3 DROYA	P13728 drosophila
10	142	16.5	1832	2 Q9S503 CRYPV	Q9s503 cryptospori
11	138	16.0	1286	2 Q9VR49 DROME	Q9vr49 drosophila
12	137.5	16.0	648	2 Q4QJ22 LEIMA	Q4qj22 leishmania
13	136.5	15.9	159	2 Q9CI73 LACLA	Q9ci73 lactococcus
14	136.5	15.9	716	2 Q4P695 USTMA	Q4p695 ustilago ma
15	136.5	15.9	1124	2 Q5CVM4 CRYPV	Q5cvm4 cryptospori
16	136.5	15.9	1126	2 Q5CIS2 CRYHO	Q5cis2 cryptospori

17	136	15.8	1937	2	Q5CV21 CRYPV	Q5cv21 cryptospori
18	135	15.7	668	2	Q4QK4 LEIMA	Q4qk4 leishmania
19	134.5	15.6	1161	1	DANA YEAST	P47179 saccharomyc
20	134	15.6	1646	2	Q5CFZ6 CRYHO	Q5cfz6 cryptospori
21	133.5	15.5	354	2	Q25402 LITS1	Q25402 litomosoid
22	133	15.4	881	2	Q8IMS9 DROME	Q8im9 drosophila
23	132.5	15.4	704	2	Q4QK1 LEIMA	Q4qk1 leishmania
24	132	15.3	519	2	Q7YTR7 CAEEL	Q7ytr7 caenorhabdi
25	130.5	15.2	753	2	Q54SK9 DICDI	Q54sk9 dictyosteli
26	128.5	14.9	916	2	Q7Y210 MONBE	Q7y210 monosiga br
27	128	14.9	333	2	Q6VAV8 ANOGE	Q6vav8 anopheles g
28	128	14.9	1763	2	Q4HX73 GIBZE	Q4hx73 gibberella
29	127	14.8	1349	2	Q8WQ4 HUMAN	Q8wq4 homo sapien
30	126.5	14.7	1117	2	Q54S26 DICDI	Q54s26 dictyosteli
31	126	14.6	457	2	Q86AK1 DICDI	Q86ak1 dictyosteli
32	126	14.6	1795	2	Q76894 DROME	Q76894 drosophila
33	125.5	14.6	605	1	WSC4 YEAST	P38739 saccharomyc
34	125	14.5	334	2	Q6VAV0 ANOGE	Q6vav0 anopheles g
35	125	14.5	1199	2	Q5CEX1 CRYHO	Q5cex1 cryptospori
36	124.5	14.5	157	2	Q8SL4 LACPL	Q8sl4 lactobacill
37	124	14.4	333	2	Q6VAV7 ANOGE	Q6vav7 anopheles g
38	124	14.4	373	2	Q76810 ANOGE	Q76810 anopheles g
39	124	14.4	510	2	Q04210 ARATH	Q04210 arabidopsis
40	123.5	14.3	374	2	Q8V0L6 9ALPH	Q8v0l6 equid herpe
41	123	14.3	334	2	Q6VAV6 ANOGE	Q6vav6 anopheles g
42	123	14.3	334	2	Q6VAV9 ANOGE	Q6vav9 anopheles g
43	123	14.3	389	2	Q8VOM0 9ALPH	Q8vom0 equid herpe
44	123	14.3	791	2	Q66VC3 9ALPH	Q66vc3 equid herpe
45	122.5	14.2	669	2	Q5ELV3 DROYA	Q5elv3 drosophila

#### ALIGNMENTS

#### RESULT 1

NTD\_LACLE  
ID NTD\_LACLE STANDARD; PRT; 156 AA.  
AC Q9R5V5;  
DT 25-OCT-2004 (Rel. 45, Created)  
DT 25-OCT-2004 (Rel. 45, Last sequence update)  
DT 01-FEB-2005 (Rel. 46, Last annotation update)  
DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-  
deoxyriboseyltransferase).  
GN Name=ntd;  
OS Lactobacillus leichmannii.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OX Lactobacillus.  
NCBI\_TaxID=28039;  
[1]  
RN  
RP NUCLEOTIDE SEQUENCE. PROTEIN SEQUENCE OF 1-25, FUNCTION, ACTIVE SITE,  
RP AND MUTAGENESIS OF GLU-97  
RX MEDLINE=95318137; PubMed=7797550; DOI=10.1074/jbc.270.26.15551;  
RA Porter D.J.T., Merrill B.M., Short S.A.;  
RT "Identification of the active site nucleophile in nucleoside 2-  
deoxyriboseyltransferase as glutamic acid 98.";  
RL J. Biol. Chem. 270:15551-15556(1995).  
[2]  
RN  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
RX MEDLINE=96419140; PubMed=8805514;  
RA Armstrong S.R., Cook W.J., Short S.A., Balick S.E.;  
RT "Crystal structures of nucleoside 2-deoxyriboseyltransferase in native  
and ligand-bound forms reveal architecture of the active site.";  
RL Structure 4:97-107(1996).  
CC  
CC -!- FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-  
deoxyribonucleosides and the transfer of the deoxyriboseyl moiety  
to an acceptor purine or pyrimidine base.  
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-riboseyl-base(1) + base(2) = 2-deoxy-  
D-riboseyl-base(2) + base(1).  
CC -!- BIOPHYSICOCHEMICAL PROPERTIES:  
CC pH dependence:  
CC Optimum pH is 6.0;  
CC -!- PATHWAY: Nucleotide metabolism.  
CC -!- SUBUNIT: Homohexamer.  
CC -!- SIMILARITY: Belongs to the nucleoside deoxyriboseyltransferase

```
CC family.
CC -1- CAUTION: Was originally (Ref.1) thought to originate from E.coli.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC PDB; 1F8X; X-ray; A/B=1-156.
CC PDB; 1F8Y; X-ray; A/B=1-156.
CC InterPro; IPR007710; N_deoxyrib_trans.
CC Pfam; PF05014; Nuc deoxyrib_tr; 1.
CC 3D-structure; Direct protein sequencing; Nucleotide metabolism;
CC Transferase.
CC INIT MET 0 0
CC ACT SITE 97 97 Nucleophile.
CC MUTAGEN 97 97 E->A: Loss of transferase activity.
CC SQ SEQUENCE 156 AA; 17949 MW; 3A3AEC3FD5B4743B CRC64;

Alignment Scores:
Pred. No.: 1,46e-68 Length: 156
Score: 843.00 Matches: 156
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.91% Indels: 0
DB: 1 Gaps: 0

US-10-049-750-13 (1-483) x NTD_LACHE (1-156)
QY 13 CCAAAAAGCAGTACTCTCGGTCGGTGTCTACTGACCGCCAAAACAAAGCCTAC 72
Db 1 ProlysylthreiletyrPheGlyAlaGlyTrpPheThrAspArgGlnAsnLysAlaTyr 20
QY 73 AAGGAAGCCATGGAGCCCTCAGGAAACCCACGATTGACCTGGAAACAGCTAGTT 132
Db 21 LysGluAlaMetGluAlaLeuLysGluAsnProthreileAspLeuGluAsnSerTyrVal 40
QY 133 CCCCTGGACCAACAGTACAAGGGTATCCGGGTGTGATCAACACCCGGAAATACCTGCATGAC 192
Db 41 ProLeuAspAsnGlnTyrLysGlyleargValAspGluHisProGluTyrLeuHisAsp 60
QY 193 AAGTTTGGGTACGGCCACTACAACACGACTTGAAACGGGATCAAGACCAACGACATC 252
Db 61 LysValTrpAlaThrAlaThrTyrAsnAspLeuAsnGlyleileysThrAsnAspIle 80
QY 253 ATGCTGGGTCTACATCCCTGACGAGAGACGTCGGCTGGCATGGAACTGGGTAC 312
Db 81 MetLeuGlyValTyrIleProAspGluGluAspValGlyLeuGlyMetGluLeuLysTyr 100
QY 313 GCCTTGAGCCAAAGCAAGTACGTCCTTTTGGTCATCCCGGACGAAGACTACGGCAAGCCG 372
Db 101 AlaLeuSerGlnGlyLysTyrValLeuLeuValIleProAspGluAspTyrGlyLysPro 120
QY 373 ATCAACTCATAGCTGGGGCGTCAGCGACAACTGATCAAGATGAGCCAGCTGAAGGAC 432
Db 121 IleAsnLeuMetSerTrpGlyValSerAspAsnValIleLysMetSerGlnLeuLysAsp 140
QY 433 TTCAACTTCAACAGCCGCTTCGACTTCTACGAAGTGGCGTATAC 480
Db 141 PheAsnPheAsnLysProArgPheAspPheTyrGluGlyAlaValTyr 156
```

```
RESULT 2
Q8RLV4_LACHE
ID Q8RLV4_LACHE PRELIMINARY; PRT; 158 AA.
AC Q8RLV4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-deoxyribosyltransferase.
GN Name:ntd;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
```

```
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CNR232;
RX MEDLINE=21964041; PubMed=11836245; DOI=10.1074/jbc.M111995200;
RA Kaminski P.A.;
RT "Functional cloning, heterologous expression, and purification of two
RT different N-deoxyribosyltransferases from Lactobacillus helveticus.";
RL J. Biol. Chem. 277:14400-14407(2002).
DR EMBL; AY064167; AAL73114.1; -; Genomic_DNA.
DR HSSP; Q9RSV5; 1F8Y.
DR SNR; Q8RLV4; 4-158.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Transferase.
SQ SEQUENCE 158 AA; 18148 MW; 00FC9E4B210C47B8 CRC64;

Alignment Scores:
Pred. No.: 5,43e-58 Length: 158
Score: 728.00 Matches: 128
Percent Similarity: 95.48% Conservative: 20
Best Local Similarity: 82.58% Mismatches: 7
Query Match: 84.55% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-13 (1-483) x Q8RLV4_LACHE (1-158)
QY 16 AAAAAAGCAGTACTCTCGTCCGCTGTTCTACTGACCGCCAAAACAAAGCCTACAAG 75
Db 4 LysLysThrLeuTyrPheGlyAlaGlyTrpPheAsnGluLysGlnAsnLysAlaTyrLys 23
QY 76 GAAGCCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAACACAGCTACGTTCCC 135
Db 24 GluAlaMetAlaAlaLeuLysGluAsnProthreileAspLeuGluAsnSerTyrValPro 43
QY 136 CTGGACCAACCCAGTACAAGGGTATCCGGGTGTGATGAACACCCGGAATACCTGCATGACAAG 195
Db 44 LeuGluAsnGlnTyrLysGlyleargValAspGluHisProGluTyrLeuHisAsnIle 63
QY 196 GTTTGGGTACGGCCACTACAACACGACTTGAAACGGGATCAAGACCAACGACATCATG 255
Db 64 GluTrpAlaSerAlaThrTyrHisAsnAspLeuValGlyleileysThrSerAspValMet 83
QY 256 CTGGGTGTCTACATCCCTGACGAGAAAGACGTCGGCTGGCATGGAACCTGGGTACGCC 315
Db 84 LeuGlyValTyrLeuProGluGluGluAspValGlyLeuGlyMetGluLeuGlyTyrAla 103
QY 316 TTGAGCCAAAGCAAGTACGTCCTTTTGGTCATCCCGGACGAAGACTACGGCAAGCCGATC 375
Db 104 LeuSerGlnGlyLysTyrIleLeuLeuValIleProAspGluAspTyrGlyLysProIle 123
QY 376 AACCTCATAGCTGGGGCGTCAGCGACCAAGCTGATCAAGATGAGCCAGCTGAAGGACTTC 435
Db 124 AsnLeuMetSerTrpGlyValCysAspAsnAlaileLysIleSerGluLeuLysAspPhe 143
QY 436 AACTTCAACAGCCGCTTCGACTTCTACGAAGTGGCGTATAC 480
Db 144 AspPheAsnLysProArgTyrAsnPheTyrAspGlyAlaValTyr 158

RESULT 3
Q5FMM8_LACAC
ID Q5FMM8_LACAC PRELIMINARY; PRT; 159 AA.
AC Q5FMM8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE 2-deoxyribosyltransferase.
GN OrderedLocusNames=LBA0145;
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
```

```

OX NCBI_TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RX PubMed=156711160; DOI=10.1073/pnas.0409188102;
RA Altmann E., Ruseell W.M., Azcarate-Peril M.A., Barrangou R.,
RA Buck B.L., McAlliffe O., Souther N., Dobson A., Duong T., Callanan M.,
RA Lick S., Hamrick A., Cano R., Klaenhammer T.R.;
RT "Complete genome sequence of the probiotic lactic acid bacterium
RT Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912(2005).
DR EMBL; CP000033; AAV42046.1; -; Genomic DNA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib tr.; 1.
DR Pfam; PF05014; N_deoxyrib tr.; 1.
KW Complete proteome; transferase.
SQ SEQUENCE 159 AA; 18252 MW; 5C7FFDBA57938FE CRC64;

Alignment Scores:
Pred. No.: 3 65e-57 Length: 159
Score: 719.00 Matches: 128
Percent Similarity: 94.81% Conservative: 18
Best Local Similarity: 83.12% Mismatches: 8
Query Match: 83.51% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-13 (1-483) x Q5FMW8_LACAC (1-159)
QY 19 AAGACGATCTACTTCGGTCCGGCTGGTTCACTGACCGCCAAACAAAGCCCTACAAGGAA 78
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6 LysThrLeuTyrPheGlyAlaGlyTyrPheAsnGluLeuGlnAsnLysAlaTyrLysAla 25
QY 79 GCATGGAGCCCTCAGGAAACCCAGCATTGACCTGGGAAACAGACTACGTTCCCTG 138
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 AlaMetGluAlaLeuLysGlnAsnProThrValAspLeuGluAsnSerTyrValProLeu 45
QY 139 GACACCCAGTACAAGGTTATCGGTTGATGACACCCGGATACCTGCATGACAGGTT 198
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46 GluAsnGlnTyrLysAspIleargValAspGluHisProGluTyrLeuHisAspIleGlu 65
QY 199 TGGGCTACGGCCACCTTACAAACACGACTTGAACGGGATCAAGCCAAACGACATCATGCTG 258
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 TrpAlaSerAlaThrTyrHisAsnAspLeuIleGlyIleLysSerSerAspIleMetLeu 85
QY 259 GTGTCTACATCCCTGACGAGACGTCGGCTGGGATGAGACTGGGTTACGCTTG 318
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
86 GlyValTyrLeuProGluGluAspValGlyLeuGlyMetGluLeuGlyTyrAlaLeu 105
QY 319 AGCCAGGCAAGTACGTCCTTTTGGTCAATCCGGACGAGACTACGCAAGCCGATCAAC 378
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 SerGlnGlyLysTyrIleLeuLeuValIleProAspGluAspTyrGlyLysProIleAsn 125
QY 379 CTCATGAGCTGGGGCTCAGGCAACAGCTGATCAAGATCAGCAGCTGAAGGACTTCAAC 438
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
126 LeuMetSerTyrGlyValCysAspAsnAlaIleLysIleSerGluLeuLysAspPheAsp 145
QY 439 TTCAACAGCCGCTTCGACTTCTACGAAGTGGCGTATAC 480
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
146 PheAsnLysProArgPheAsnPheTyrAspGlyAlaValTyr 159

RESULT 4
NTD_LACJO
ID NTD_LACJO STANDARD; PRT; 149 AA.
AC Q74LQ9;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-
DE deoxyribosyltransferase).
GN Name=ntd; OrderedLocusNames=LJ0124;
OS Lactobacillus johnsonii.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
```

```
Db 142 AsnPheTyrAspGlyAlaValTyr 149
:::|||||:::|||||
RESULT 5
Q9KWF0 LACHE
ID Q9KWF0 LACHE PRELIMINARY; PRT; 158 AA.
AC Q9KWF0
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleoside deoxyribosyltransferase-II.
GN Name=ndtB;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21012342; PubMed=11129605; DOI=10.1271/bbb.64.2243;
RA Okuyama K., Nouguchi T.;
RT "Molecular cloning and expression of the nucleoside
RT deoxyribosyltransferase-II gene from Lactobacillus helveticus.";
RL Biosci. Biotechnol. Biochem. 64:2243-2245(2000).
DR EMBL; AB039914; BAA92683.2; -; Genomic_DNA.
DR PIR; J07522; J07522.
DR HSP; Q9RSV5; I78Y.
DR SMR; Q9KWF0; 4-158.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N deoxyrib trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Transferase.
SQ SEQUENCE 158 AA; 18318 MW; D661273AD89309CD CRC64;

Alignment Scores:
Pred. No.: 2,37e-49 Length: 158
Score: 634.00 Matches: 113
Percent Similarity: 87.10% Conservative: 22
Best Local Similarity: 72.90% Mismatches: 20
Query Match: 73.64% Indels: 0
DB: 2 Gaps: 0

US-10-049-750-13 (1-483) x Q9KWF0_LACHE (1-158)
Qy 16 AAAAGACGATCTACTTCGGTGGCGGTGTTCACTGACCGCCAAACAAAGCCTACAAG 75
|||||:::|||||:::|||||
Db 4 LysLysThrLeuTyrPheGlyAlaGlyTrpPheAsnGluLysGlnAsnLysAlaTyrLys 23
|||||
Qy 76 GAAGCCATGGAGCCCTCAGGAAACCCACGATTGACCTGGGAAACAGCTAGCTTCCC 135
|||||
Db 24 GluAlaMetAlaAlaLeuLysGluAsnProThrValAspLeuGluAsnSerTyrValPro 43
|||||
Qy 136 CTGGACAACAGTACAAGGGTATCCGGTGTGATCAACACCGGAATACCTGCATGACAAG 195
|||||
Db 44 LeuGluAsnGlnTyrLysGlyIleArgIleAspGluHisProGlnTyrLeuHisAsnIle 63
|||||
Qy 196 GTTTGGGCTACGGCCACTTACAACACGACTTGTAACGGGATCAAGACCAACGACATCATG 255
|||||
Db 64 GluTrpAlaSerAlaThrTyrHisAsnAspLeuValGlyIleLysThrSerAspValLeu 83
|||||
Qy 256 CTGGGTGTATACCTCCCTGACGAGAGAGAGCTGGCGCTGGGCATGGAACTGGGTAGCC 315
|||||
Db 84 LeuGlyValTyrLeuProGlnGluHisValGlyLeuGlyMetGluLeuGlyTyrPro 103
|||||
Qy 316 TTGAGCAAGCAAGTAGTACGTCCTTTTGTCTATCCGGACGAAGACTACGGCAAGCCCATC 375
|||||
Db 104 LeuSerGlnGlyLysLeuPhePheTrpPheSerHisMetLysAspTyrGlyLysProIle 123
|||||
Qy 376 AACCTCATGAGCTGGGGCGTCAGCGACACCAACGTGATCAAGATGAGCCAGCTGAAGGACTTC 435
|||||
Db 124 IleLeuMetSerTrpGlyValCysAspAsnAlaSerGlnIleSerGluLeuLysAspPhe 143
|||||
Qy 436 AACTTCAACAGCGCGCTCGACTTCATTACGAAGTCCGCTATAC 480
:::|||||:::|||||
```

```
Db 144 AspPheAsnLysProArgTyrAsnPheTyrAspGlyAlaValTyr 158
NTD_LACFE
ID NTD_LACFE STANDARD; PRT; 168 AA.
AC Q6YNI5;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-
DE deoxyribosyltransferase).
GN Name=ntd;
OS Lactobacillus fermentum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1613;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=CIP 102980T;
RA Kaminski P.A.;
RT "Characterization of the Lactobacillus fermentum N-
RT deoxyribosyltransferase.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-
CC deoxyribonucleosides and the transfer of the deoxyribosyl moiety
CC to an acceptor purine or pyrimidine base (By similarity).
CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose(1)-base(2) = 2-deoxy-
CC D-ribose(1)-base(2) + base(1).
CC -!- PATHWAY: Nucleotide metabolism.
CC -!- SIMILARITY: Belongs to the nucleoside deoxyribosyltransferase
CC family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; AY064168; AAL73115.1; -; Genomic_DNA.
CC InterPro; IPR007710; N deoxyrib trans.
CC Pfam; PF05014; Nuc_deoxyrib_tr; 1.
CC Nucleotide metabolism; Transferase.
FT ACT_SITE 103 103 Nucleophile (By similarity).
SQ SEQUENCE 168 AA; 18896 MW; 362162F43586C317 CRC64;

Alignment Scores:
Pred. No.: 2,73e-10 Length: 168
Score: 209.00 Matches: 55
Percent Similarity: 51.66% Conservative: 23
Best Local Similarity: 36.42% Mismatches: 59
Query Match: 24.27% Indels: 14
DB: 1 Gaps: 7

US-10-049-750-13 (1-483) x NTD_LACFE (1-168)
Qy 25 ATCTACTTCGGTCCGCTGGTTCACCTGACCGCCAAACAAAGCCTACAAGGAGCCATG 84
|||||
Db 12 IleTyrLeuAlaThrSerPhePheAsnGluGluGlnArgAlaArgIleProGlnAlaLeu 31
|||||
Qy 85 GAAGCCCTCAAGAAACCCACGATTGACCTGGAAACAGCTACGTTCCCTGGACAAC 144
|||||
Db 32 AlaGlnLeuGluAlaAsnProThrValGlyValHisGln-----ProPheAspPhe 49
|||||
Qy 145 CAGTACAAGGTATCCGGGTGATGAACACCGGAATACCTGCATGACAAGGTT---TGG 201
|||||
Db 50 GlnTyrIleAspAlaArgValAspSerAspProAlaGlyValPheGlySerLeuGluTrp 69
|||||
Qy 202 GCTACGGCCACCTACAACACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGGT 261
|||||
Db 70 GlnIleAlaThrTyrAsnAsnAspLeuAsnAlaValGlyThrSerAspValCysValAla 89
|||||
Qy 262 GTCTACATCCCTGACGAGAGAGCGTGGCGCTGGGATGGAACCTGGGTACGCTTGAGC 321
:::|||||:::|||||
```

```

Db 90 LeuTyrAspMetAspGlnIleAspGluGlyIleCysMetGluIleGlyMetPheValAla 109
Qy 322 CAAGCAAG---TAGTCTCTTTTGTATCCCGACGAGACTACGGCAAG---CCGATC 375
Db 110 LeuHisLysProIleValLeuLeuProPheThrLysLysAspLysSerAlaTyrGluAla 129
Qy 376 AACCTCATG-----AGCTGGGGCGTCAGCGACAGCGTATCAAGATG 417
Db 130 AsnLeuMetLeuAlaArgGlyValThrTrp---LeuGluProAsn-----AspPhe 146
Qy 418 AGCCAGCTGAAGGACTTCAACTCAACCAAGCGC 450
Db 147 SerProLeuLysAspPheAsnHisPro 157

RESULT 7
Q8RLY5_LACHE
ID Q8RLY5_LACHE PRELIMINARY; PRT; 167 AA.
AC Q8RLY5
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Purine trans deoxyribosylase (Nucleoside deoxyribosyltransferase-I)
DE (EC 2.4.2.6).
GN Name=ptd; Synonyms=ndtA;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CNR232;
RX MEDLINE=21964041; PubMed=11836245; DOI=10.1074/jbc.M111995200;
RA Kaminski P.A.;
RT "Functional cloning, heterologous expression, and purification of two
different N-deoxyribosyltransferases from Lactobacillus helveticus.";
RL J. Biol. Chem. 277:14400-14407(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Okuyama K., Noguchi T.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY064166; AAL73113.1; -; Genomic DNA.
DR EMBL; AB076265; BAC00952.1; -; Genomic DNA.
DR FDB; 182D; X-ray; A/B/C-1-167.
DR PDB; 182G; X-ray; A/B/C-1-167.
DR PDB; 182I; X-ray; A/B/C-1-167.
DR PDB; 182L; X-ray; A/B/C-1-167.
DR PDB; 183F; X-ray; A/B/C-1-167.
DR GO; GO:0050144; P:nucleoside deoxyribosyltransferase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 167 AA; 18713 MW; 5540581511CB4787 CRC64;

```

Alignment Scores:

Pred. No.:	7.64e-07	Length:	167
Score:	171.50	Matches:	49
Percent Similarity:	53.21%	Conservative:	34
Best Local Similarity:	31.41%	Mismatches:	58
Query Match:	19.92%	Indels:	15
DB:	2	Gaps:	7

```

US-10-049-750-13 (1-483) x Q8RLY5_LACHE (1-167)
Qy 10 ATGCCAAAAGACCATCTACTTCGGTCCGGCTGTTCACTGACGCCCAAAACAAAGCC 69
Db 5 ValProThrGlyLysIleValLeuGlySerProPheThrLysSerAlaGlnArgGluArg 24
Qy 70 TACAAGGAAGCATGGAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAACAGCTAC 129
Db 25 AlaAlaLysAlaLysGluLeuLeuAlaLysAsnProSerIle-----AlaHisValPhe 42
Qy 130 GTTCCCTCGGCAACACGATCAAGGGGTATCCGGGTGTGATGAACCCCGAATAC----- 183

```

```

Db 43 PheProPheAspGlyPheThrAsp---ProAspGluLysAsnProGluIleGlyGly 61
Qy 184 CTGCATGACAGGTTTGGGCTACGGCCACCTACCAACACGACTTGAACGGGATCAAGACC 243
Db 62 IleArgSerMetValTrpArgAspAlaThrTyrGlnAsnAspLeuThrGlyIleSerAsn 81
Qy 244 AACGACATCATGCTGGGTGCTACATCCCTCAGCAAGAAGACGTCGGCCTGGGCATGAA 303
Db 82 AlaThrCysGlyValPheLeuTyrAspMetAspGlnLeuAspAspGlySerAlaPheGlu 101
Qy 304 CTGGTTTACGGCTTGAGCAAGCAAGTACGTCTCTTTTGGTC-----ATCCCG 351
Db 102 IleGlyPheMetArgAlaMetHisLysProValIleLeuValProPheThrGluHisPro 121
Qy 352 GACGAAGACTACGCAAGCCGATCAACTCATG---AGCTGGGGCGTCAGC----- 399
Db 122 GluLysGlu-----LysLysMetAsnLeuMetIleAlaGlnGlyValThrThrIleIle 139
Qy 400 GACAACGTGATCAAGATGAGCCAGCTGACGAGCTCAAGACTTCAACTCAACAAAG 447
Db 140 AspGlyAsnThrGluPheGluLysLeuAlaAspTyrAsnPheAsnGlu 155

RESULT 8
Q5FINO_LACAC
ID Q5FINO_LACAC PRELIMINARY; PRT; 167 AA.
AC Q5FINO;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Putative deoxyribosyltransferase.
DE OrderedLocusNames=LBA1631;
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RX PubMed=15671160; DOI=10.1073/pnas.0409188102;
RA Altermann E., Russell W.M., Azcarate-Peril M.A., Barrangou R.,
Buck B.L., McAuliffe O., Souther N., Dobson A., Duong T., Callanan M.,
Lick S., Hamrick A., Cano R., Klaenhammer T.R.;
RT "Complete genome sequence of the probiotic lactic acid bacterium
Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912(2005).
DR EMBL; CP000033; AAV43444.1; -; Genomic DNA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 167 AA; 18933 MW; A80408F06C3E4D98 CRC64;

```

Alignment Scores:

Pred. No.:	2.79e-05	Length:	167
Score:	154.50	Matches:	49
Percent Similarity:	49.01%	Conservative:	25
Best Local Similarity:	32.45%	Mismatches:	60
Query Match:	17.94%	Indels:	17
DB:	2	Gaps:	7

```

US-10-049-750-13 (1-483) x Q5FINO_LACAC (1-167)
Qy 25 ATCTACTTCGGTGGCGGCTGTTCTACTGACCGCCAAAACAAAGCCTACAAGGAAGCCATG 84
Db 10 IleTyrLeuGlyThrProPheThrAsnAspGlnArgAlaArgValLysLysAlaArg 29
Qy 85 GAAGCCCTCAAGGAAACCCCAACGATTGACCTGGAAACACAGCTAGTCCCTCGGACAAAC 144
Db 30 AlaLeuLeuGluGlnAsnProThrVal-----ValArgValHisPheProPheAspGln 47
Qy 145 CAGTACACAGGGTATCCGGGTGAT-----GAACACCCGGAATAC-----CTGCAT 189

```











```

OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Iowa type II;
RX PubMed=15044751; DOI=10.1126/science.1094786;
RA Abrahamson M.S., Templeton T.J., Enomoto S., Abrahamante J.E., Zhu G.,
RA Lantto C.A., Deng M., Liu C., Widmer G., Tzipori S., Buck G.A., Xu P.,
RA Bankier A.T., Dear P.H., Konfortov B.A., Spriggs H.F., Iyer L.,
RA Anantharaman V., Aravind L., Kapur V.;
RT "Complete genome sequence of the apicomplexan, Cryptosporidium
RT parvum.";
RL Science 304:441-445(2004).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAEE01000003; EAK89514.1; -; Genomic_DNA.
DR InterPro; IPR009081; ACP_like.
DR InterPro; IPR000074; ApoA1_A4_E.
DR InterPro; IPR010980; Cyt_c_b562.
DR InterPro; IPR011992; EF-Hand type.
DR InterPro; IPR002563; Flv_red_FMN_bind.
DR InterPro; IPR004281; IL12.
DR InterPro; IPR010503; LT-IIIB.
DR InterPro; IPR000429; Prot_inh_hirudin.
DR InterPro; IPR000783; RNAPol_RPB5.
DR InterPro; IPR009032; Ypu_cyt.
SQ SEQUENCE 1124 AA; 120832 MW; E3EB73E8DC34C87E CRC64;

Alignment Scores:
Pred. No.: 0.00162 Length: 1124
Score: 136.50 Matches: 46
Percent Similarity: 50.76% Conservative: 21
Best Local Similarity: 34.85% Mismatches: 52
Query Match: 15.85% Indels: 13
DB: 2 Gaps: 4

US-10-049-750-13 (1-483) x Q5CVN4_CRYPV (1-1124)
QY 86 AGCCCTCAAGAAACCCACGATTGACCTGGAAACAGCTAGCTCCCTGGACAACC 145
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 870 LysGlySerGluCysValArgProSerThrThrThrThrThrThrThrThrThrThr 889
QY 146 AGTACAAGGGTATCCGGGTGATGAACACCCGGAATACCTGCATGACAAAGTTTGGGCTA 205
DB ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 890 ThrThrArg--SerThrThrThrThrThrThrThrThrThrThrThrThrArg----- 905
QY 206 CGGCCACTACAACACGACTTGAACGGGATCAAGACCAACGACATCATGCTGGGTGCT 265
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 906 -----ProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 923
QY 266 ACATCCCTGACGAGAGACGTCGGCTGGGCATGGGAACCTGGGTAGCCTTGAGCCAAG 325
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 924 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 943
QY 326 GCAAGTACGTCCTTTTGGTCATCCCGGACGACGAGGATACGGCAAGCCGATCAACCTCATGA 385
DB ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 944 ProThrThrThr-----ThrThrArgProThrThrThrThrThrThrThrThrThr 961
QY 386 GCTGGGGCGTCAGCGACAACGTGATCAAGATGAGCCAGCTGAAGACCTCAACTCAACA 445
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 962 ThrThrArgProThrThrThr-----ThrThrArgProThrThrThrThrThrThr 976
QY 446 AGCGCGCTTCGACTTCTACGAAGGTGCGGTACT 481
DB ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 977 ThrArgProThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 988

Search completed: December 23, 2005, 22:42:39
Job time : 137.382 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:47:05 ; Search time 188 Seconds

(without alignments)  
366.928 Million cell updates/sec

Title: US-10-049-750-14

Perfect score: 848

Sequence: 1 MPKXTIYFGAGWFTDRQNK.....QLKDFNFKNKPRDFYEGAVY 157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A. Geneseq 21.\*

- 1: geneseqp1980s.\*
- 2: geneseqp1990s.\*
- 3: geneseqp2000s.\*
- 4: geneseqp2001s.\*
- 5: geneseqp2002s.\*
- 6: geneseqp2003as.\*
- 7: geneseqp2003bs.\*
- 8: geneseqp2004s.\*
- 9: geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	157	4	AAB67589 Amino aci
2	842	99.3	157	8	Adr97103 N-deoxyri
3	728	85.8	158	6	Abp98175 Amino aci
4	719	84.8	159	6	Abp98180 Amino aci
5	637	75.1	158	5	Abg66170 L. helvet
6	629	74.2	133	6	Abp98179 Amino aci
7	405	47.8	84	6	Abp98178 Amino aci
8	209	24.6	168	6	Abp98177 Amino aci
9	166.5	19.6	167	6	Abp98176 Amino aci
10	136.5	16.1	159	5	Abb53802 Lactococc
11	96	11.3	1149	5	Abb77620 AnEPV P4a
12	94	11.1	750	1	Aap82940 Atrial na
13	91.5	10.8	156	7	Adc96860 E. faeciu
14	91	10.7	719	9	Adw47705 F. magna
15	86.5	10.2	462	5	Abb54050 Lactococc
16	85.5	10.1	438	4	Aab95642 Human pro
17	85.5	10.1	456	8	Adh41469 Novel hum
18	85.5	10.1	464	3	Adb23624 Human sec
19	85.5	10.1	487	8	Adh41499 Novel hum
20	85.5	10.1	487	8	Adh41497 Novel hum
21	85.5	10.1	487	8	Adh41493 Novel hum
22	85.5	10.1	487	8	Adh41495 Novel hum
23	85.5	10.1	487	8	Adh41463 Novel hum
24	85.5	10.1	487	8	AdS88243 Human pro

25	84.5	10.0	243	8	ADH41485	Novel hum
26	84.5	10.0	247	8	ADH41467	Novel hum
27	84.5	10.0	396	4	AAB90548	Human sec
28	84.5	10.0	396	5	ABG65427	Human alb
29	84.5	10.0	396	8	ADL78694	Albumin f
30	84.5	10.0	452	8	ADH41489	Novel hum
31	84.5	10.0	456	8	ADH41471	Novel hum
32	84.5	10.0	456	8	ADH41465	Novel hum
33	84.5	10.0	456	8	ADH41477	Novel hum
34	84.5	10.0	456	8	ADH41479	Novel hum
35	84.5	10.0	487	4	AAU08684	Human FCT
36	84.5	10.0	487	4	AAU08683	Human FCT
37	84.5	10.0	487	7	ADB32033	Human FCT
38	84.5	10.0	487	7	ADB32035	Human FCT
39	84.5	10.0	487	8	ADH41491	Novel hum
40	84.5	10.0	493	8	ADH41473	Novel hum
41	84	9.9	706	6	ABO53100	Human put
42	84	9.9	1042	9	ADX07993	Cyclin-de
43	84	9.9	1046	7	ADE58906	Human Pro
44	84	9.9	1046	7	ADD45534	Human Pro
45	84	9.9	1046	7	ADE61258	Human Pro

#### ALIGNMENTS

#### RESULT 1

AAB67589

ID AAB67589 standard; protein; 157 AA.

AC AAB67589;

DT 29-MAY-2001 (first entry)

XX

DE Amino acid sequence of a nucleoside 2-deoxyribosyltransferase.

KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;

KW purine nucleoside phosphorylase; phosphopentose mutase;

KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;

KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase.

XX

OS Lactobacillus leichmannii.

XX WO200114566-A2.

XX 01-MAR-2001.

XX 18-AUG-2000; 2000WO-EP008088.

XX 20-AUG-1999; 99EP-00116425.

XX (HOFF ) ROCHE DIAGNOSTICS GMBH.

XX (INSP ) INST PASTEUR.

XX (PHAR-) PHARMA-WALDHOF GMBH & CO KG.

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;

PI Marliere P, Pochet S;

DR WPI; 2001-235026/24.

DR N-PSDB; AAF55445.

PT In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside and an inorganic phosphate.

PS Disclosure; Page 62-63; 73pp; English.

XX The present sequence represents a nucleoside 2-deoxyribosyltransferase enzyme. This enzyme is involved in the biosynthesis of

CC deoxyribonucleosides, and is used in the method of the invention. The

CC specification describes a method for the in vitro enzymatic synthesis of

CC deoxyribonucleosides. The method comprises reacting deoxyribose 1-

CC phosphate and a nucleobase to form a deoxyribonucleoside and an inorganic

CC phosphate. Enzymes which may be used in the method of the invention  
 CC include thymidine phosphorylase, purine nucleoside phosphorylase,  
 CC phosphotriose mutase, phosphopentose aldolase, fructose 1,6-diphosphate  
 CC aldolase, deoxyribokinase, and nucleoside 2-deoxyribosyltransferase  
 XX  
 SQ Sequence 157 AA;

Query Match 100.0%; Score 848; DB 4; Length 157;  
 Best Local Similarity 100.0%; Pred. NO. 5.8e-84;  
 Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPKKTIFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGRVDEHPEYLH 60  
 DB 1 MPKKTIFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGRVDEHPEYLH 60  
 QY 61 DKWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSOGKYVLLVIPDEDYDK 120  
 DB 61 DKWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSOGKYVLLVIPDEDYDK 120  
 QY 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157  
 DB 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157

RESULT 2  
 ADR97103  
 ID ADR97103 standard; protein; 157 AA.  
 XX  
 AC ADR97103;  
 XX  
 DT 16-DEC-2004 (first entry)  
 DE N-deoxyribosyltransferase (DTP) with G9S mutant, seq id 1.  
 XX  
 KW Cytostatic; antimicrobial; N-deoxyribosyltransferase; DDTP;  
 KW N-deoxyribosyltransferase; DTP; antitumour; infection; ddc; ddi; mutein.  
 XX  
 OS Lactobacillus leichmannii.  
 XX  
 OS Synthetic.  
 PH Key Location/Qualifiers  
 FT Misc-difference 9 /note= "Wild-type Gly replaced with Ser"  
 FT  
 XX  
 XX FR2852968-A1.  
 XX  
 XX 01-OCT-2004.  
 XX  
 XX 28-MAR-2003; 2003FR-00003910.  
 XX  
 XX 28-MAR-2003; 2003FR-00003910.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Kaminski PA;  
 XX  
 XX WPI; 2004-693069/68.  
 XX N-PSDB; ADR97104.  
 XX  
 XX Method for evolution of proteins to modify characteristics, useful  
 PT particularly for evolving an N-deoxyribosyltransferase to produce an N-  
 PT dideoxyribosyltransferase for preparation of antitumor nucleosides and  
 PT nucleotides.  
 XX  
 XX Claim 15; SEQ ID NO 2; 51pp; French.  
 XX  
 XX The invention relates to a method for evolution of a protein (X) so as to  
 CC modify its characteristics. Further disclosed are mutated proteins (X\*)  
 CC produced by the new method and with altered activity, relative to (X).  
 CC Also disclosed is a nucleic acid (I) comprising a sequence for N-  
 CC dideoxyribosyltransferase (DDTP), obtained from an N-  
 CC deoxyribosyltransferase (DTP) by using the new method. The method is

CC specifically used to evolve an N-deoxyribosyltransferase (DTP) to an N-  
 CC dideoxyribosyltransferase (DDTP), i.e. an enzyme able to transfer  
 CC dideoxyribose from one dideoxyribonucleotide to another. Such an enzyme  
 CC is used for preparation of nucleotides and nucleosides that have  
 CC antitumour activity or can be used to treat infections, particularly ddc  
 CC and ddi. The current sequence represents the N-deoxyribosyltransferase  
 CC (DTP) amino acid sequence containing a G9S mutation. This protein  
 XX displays the activity of N-dideoxyribosyltransferase.  
 SQ Sequence 157 AA;

Query Match 99.3%; Score 842; DB 8; Length 157;  
 Best Local Similarity 99.4%; Pred. NO. 2.6e-83;  
 Matches 156; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPKKTIFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGRVDEHPEYLH 60  
 DB 1 MPKKTIFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGRVDEHPEYLH 60  
 QY 61 DKWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSOGKYVLLVIPDEDYDK 120  
 DB 61 DKWATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMEIGYALSOGKYVLLVIPDEDYDK 120  
 QY 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157  
 DB 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157

RESULT 3  
 ABP98175  
 ID ABP98175 standard; protein; 158 AA.  
 XX  
 AC ABP98175;  
 XX  
 DT 11-AUG-2003 (first entry)  
 DE Amino acid sequence of N-deoxyribosyltransferase protein.  
 XX  
 KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;  
 KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.  
 XX  
 OS Lactobacillus helveticus.  
 XX  
 XX WO2003025163-A2.  
 XX  
 XX 27-MAR-2003.  
 XX  
 XX 12-SEP-2002; 2002WO-FR003120.  
 XX  
 XX 14-SEP-2001; 2001FR-00011911.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX (NARE-) INST NAT RECH AGRONOMIQUE.  
 XX  
 XX Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;  
 XX WPI; 2003-313351/30.  
 XX N-PSDB; ACC43580, ACC43586.  
 XX  
 XX New N-deoxyribosyltransferase from Lactobacillus, useful for preparing  
 PT deoxyribonucleotides, e.g. for use as antiviral agents, also related  
 PT nucleic acid and antibodies.  
 XX  
 XX Claim 1; Page 61; 70pp; French.  
 XX  
 XX The present sequence represents a N-deoxyribosyltransferase protein from  
 CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-  
 CC deoxyribosyltransferase protein is useful to raise specific antibodies,  
 CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,  
 CC particularly those containing non-natural bases. These  
 CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),  
 CC parasitic or fungal infections, for antitumour chemotherapy, and as  
 CC insecticides or herbicides

```

XX SQ Sequence 158 AA;
Query Match 85.8%; Score 728; DB 6; Length 158;
Best Local Similarity 82.6%; Pred. No. 7.3e-71;
Matches 128; Conservative 20; Mismatches 7; Indels 0; Gaps 0;

Qy 3 KKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPEYLDHK 62
Db 4 KKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPEYLDHK 63
Qy 63 VWATATYNDNLNGIKTNDIMLGVIIPDEEDVGLGMELGALYSQGYVLLVIPDDEYDKPI 122
Db 64 EWASATYHNDLVGIKTSDVLMGVLPDEEDVGLGMELGALYSQGYVLLVIPDDEYDKPI 123
Qy 123 NLMGVSNDNVIKMSQLKDFNFKPRDFYEGAVY 157
Db 124 NLMGVSNDNAIKISELKDFNFKPRDFYEGAVY 158

RESULT 4
ID ABP98180 standard; protein; 159 AA.
XX AC ABP98180;
XX DT 11-AUG-2003 (first entry)
XX DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
XX OS deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX PN Lactobacillus acidophilus.
XX WO2003025163-A2.
XX PD 27-MAR-2003.
XX PF 12-SEP-2002; 2002WO-FR003120.
XX PR 14-SEP-2001; 2001FR-00011911.
XX PA (INSP ) INST PASTEUR.
XX PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX WPI; 2003-313351/30.
XX DR N-PSDB; ACC43585.
XX PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
XX PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
XX PS nucleic acid and antibodies.
XX Claim 1; Page 68; 70pp; French.
XX The present sequence represents a N-deoxyribosyltransferase protein from
XX CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
XX CC deoxyribosyltransferase protein is useful to raise specific antibodies,
XX CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
XX CC particularly those containing non-natural bases. These
XX CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
XX CC parasitic or fungal infections, for antitumour chemotherapy, and as
XX CC insecticides or herbicides
XX SQ Sequence 159 AA;
Query Match 84.8%; Score 719; DB 6; Length 159;
Best Local Similarity 83.1%; Pred. No. 7e-70;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

Qy 4 KTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPEYLDHKV 63

```

---

```

Db 6 KTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLENQYKDIRVDEHPEYLDHK 65
Qy 64 WATATYNDNLNGIKTNDIMLGVIIPDEEDVGLGMELGALYSQGYVLLVIPDDEYDKPI 123
Db 66 WASATYHNDLVGIKTSDVLMGVLPDEEDVGLGMELGALYSQGYVLLVIPDDEYDKPI 125
Qy 124 LMSGVSNDNVIKMSQLKDFNFKPRDFYEGAVY 157
Db 126 LMSGVSNDNAIKISELKDFNFKPRDFYEGAVY 159

RESULT 5
ID AAG66170 standard; protein; 158 AA.
XX AC AAG66170;
XX DT 17-JUN-2002 (first entry)
XX DE L. helveticus nucleoside deoxyribosyltransferase-II.
XX KW Nucleoside deoxyribosyltransferase-II; enzyme; deoxynucleoside; drug;
XX OS recombinant; ntdB.
XX PN Lactobacillus helveticus.
XX JP2002051781-A.
XX PD 19-FEB-2002.
XX PF 08-AUG-2000; 2000JP-00239443.
XX PR 08-AUG-2000; 2000JP-00239443.
XX PA (YAMS ) YAMASA SHOYU KK.
XX WPI; 2002-299403/34.
XX DR N-PSDB; ABL40579.
XX PT Enzymatic preparation of deoxynucleoside.
XX PS Claim 1; Page 10; 14pp; Japanese.
XX CC The present invention relates to a recombinant nucleoside
XX CC deoxyribosyltransferase-II. The nucleoside deoxyribosyltransferase-II and
XX CC the encoding gene are useful in an enzymatic preparation of a
XX CC deoxynucleoside. The deoxynucleoside is useful as the raw material for
XX CC various drugs. The present sequence represents the L. helveticus
XX CC recombinant nucleoside deoxyribosyltransferase-II (ntdB)
XX SQ Sequence 158 AA;
Query Match 75.1%; Score 637; DB 5; Length 158;
Best Local Similarity 73.5%; Pred. No. 6.2e-61;
Matches 114; Conservative 21; Mismatches 20; Indels 0; Gaps 0;

Qy 3 KKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPEYLDHK 62
Db 4 KKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSVPLDNOYKGI RVDHEPEYLDHK 63
Qy 63 VWATATYNDNLNGIKTNDIMLGVIIPDEEDVGLGMELGALYSQGYVLLVIPDDEYDKPI 122
Db 64 EWASATYHNDLVGIKTSDVLMGVLPDEEDVGLGMELGALYSQGYVLLVIPDDEYDKPI 123
Qy 123 NLMGVSNDNVIKMSQLKDFNFKPRDFYEGAVY 157
Db 124 ILMGVSNDNAISQISELKDFNFKPRDFYEGAVY 158

RESULT 6
ID ABP98179 standard; protein; 133 AA.

```

```

XX AC ABP98179;
XX DT .11-AUG-2003 (first entry)
XX DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
XX KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX OS Lactobacillus amylovorus.
XX XX WO2003025163-A2.
XX PD 27-MAR-2003.
XX PF 12-SEP-2002; 2002WO-FR003120.
XX PR 14-SEP-2001; 2001FR-00011911.
XX XX (INSP ) INST PASTEUR.
XX PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX WPI; 2003-313351/30.
XX DR N-PSDB; ACC43584.
XX PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
XX PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
XX PS nucleic acid and antibodies.
XX PS Claim 1; Page 66-67; 70pp; French.
XX CC The present sequence represents a N-deoxyribosyltransferase protein from
XX CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
XX CC deoxyribosyltransferase protein is useful to raise specific antibodies,
XX CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
XX CC particularly those containing non-natural bases. These
XX CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
XX CC parasitic or fungal infections, for antitumour chemotherapy, and as
XX CC insecticides or herbicides
XX SQ Sequence 133 AA;
XX Query Match 74.2%; Score 629; DB 6; Length 133;
XX Best Local Similarity 84.2%; Pred. No. 3.6e-60;
XX Matches 112; Conservative 17; Mismatches 4; Indels 0; Gaps 0;
QY 25 MEALKENPTIDLENSYVPLDNQYKGIKRVDEHPHYLHDKVWATATYNNDLNGIKTNDIMLG 84
DB 1 MEALKKNPTVDLENSYVPLDNQYKGIKRVDEHPHYLHDKVWATATYNNDLNGIKSSDVMLG 60
QY 85 VYLPDEEDVGLGMLGVALSGQKVLVLPDEDYGKPINLMSGVSNDVIMQSLKDFNF 144
DB 61 VYLPDEEDVGLGMLGVALSGQKVLVLPDEDYGKPINLMSGVCNDVIKISLKDQDF 120
QY 145 NKPRFDYEGAVY 157
DB 121 NRPRNFYDGAVY 133
RESULT 7
ABP98178
ID ABP98178 standard; protein; 84 AA.
XX AC ABP98178;
XX DT .11-AUG-2003 (first entry)
XX DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;

```

```

KW XX deoxyribonucleotide synthesis; infection; insecticide; herbicide.
OS XX Lactobacillus crispatus.
PN XX WO2003025163-A2.
XX PD 27-MAR-2003.
XX PF 12-SEP-2002; 2002WO-FR003120.
XX PR 14-SEP-2001; 2001FR-00011911.
XX XX (INSP ) INST PASTEUR.
XX PA (NARE-) INST NAT RECH AGRONOMIQUE.
XX PI Kaminski P, Tailliez P, Marliere P, Quenee P, Cotaya R;
XX WPI; 2003-313351/30.
XX DR N-PSDB; ACC43583.
XX PT New N-deoxyribosyltransferase from Lactobacillus, useful for preparing
XX PT deoxyribonucleotides, e.g. for use as antiviral agents, also related
XX PS nucleic acid and antibodies.
XX PS Claim 1; Page 65; 70pp; French.
XX CC The present sequence represents a N-deoxyribosyltransferase protein from
XX CC Lactobacillus. The enzyme is encoded by the ntd gene. The N-
XX CC deoxyribosyltransferase protein is useful to raise specific antibodies,
XX CC and for in vitro or in vivo enzymatic synthesis of deoxyribonucleotides,
XX CC particularly those containing non-natural bases. These
XX CC deoxyribonucleotides are useful for treating bacterial, viral (e.g. HIV),
XX CC parasitic or fungal infections, for antitumour chemotherapy, and as
XX CC insecticides or herbicides
XX SQ Sequence 84 AA;
XX Query Match 47.8%; Score 405; DB 6; Length 84;
XX Best Local Similarity 86.9%; Pred. No. 5.1e-36;
XX Matches 73; Conservative 8; Mismatches 3; Indels 0; Gaps 0;
QY 45 NOYKGIKRVDEHPHYLHDKVWATATYNNDLNGIKTNDIMLGVIYPDEDYGLGMLGVALS 104
DB 1 NOYKGIKRVDEHPHYLHDKVWATATYNNDLNGIKSSDMLGVLPEEDVGLGMLGVALS 60
QY 105 QGKYVLLVLPDEDYGKPINLMSWG 128
DB 61 KGKYLVLVLPDEDYGKPINLMSWG 84
RESULT 8
ABP98177
ID ABP98177 standard; protein; 168 AA.
XX AC ABP98177;
XX DT .11-AUG-2003 (first entry)
XX DE Amino acid sequence of N-deoxyribosyltransferase protein.
XX KW N-deoxyribosyltransferase; ntd gene; enzyme; antitumour chemotherapy;
XX KW deoxyribonucleotide synthesis; infection; insecticide; herbicide.
XX OS Lactobacillus fermentum.
XX PN WO2003025163-A2.
XX PD 27-MAR-2003.
XX PF 12-SEP-2002; 2002WO-FR003120.
XX PR 14-SEP-2001; 2001FR-00011911.
XX

```







Query Match 11.1%; Score 94; DB 1; Length 750;  
Best Local Similarity 26.5%; Pred. No. 1;  
Matches 48; Conservative 21; Mismatches 48; Indels 64; Gaps 10;  
1 MPKKTIFYGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEH----- 55  
226 LPSRDYECTG-----IYKACTAY-----VDFMISVAKLRQKGLPVDENQLSLE 272  
56 -----PYLHDKWATATYNN-----DLNGIK-----TNDIM- 82  
273 MNKMELEKEITANATKSEDRNDPMLLYNKMTLAQIQNNFSLINGKFSFWSNFTNEIMS 332  
83 -LGVVIPPEDVGLGMELGYA---LSQKYVLLVLPDEDYKPINLMSWG-VSDNVIKMS 137  
333 TVNINIPNEEDV-----VYAYPEYLTKLKILITKYSARDLQ---NLMSWRFMIDLVSSLS 384  
138 Q 138  
385 R 385

RESULT 13  
ADC96860  
ID ADC96860 standard; protein; 156 AA.  
XX AC ADC96860;  
XX DT 01-JAN-2004 (first entry)  
XX DE E. faecium protein sequence SEQ ID 6487.  
XX KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
XX KW abdominal-pelvic infection.  
XX OS Enterococcus faecium.  
XX US6583275-B1.  
XX PN 24-JUN-2003.  
XX PD 30-JUN-1998; 98US-00107532.  
XX PF 02-JUL-1997; 97US-0051571P.  
XX PR 14-MAY-1998; 98US-0085598P.  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX PI Doucette-Stamm LA, Bush D;  
XX DR N-PSDB; ADC93206.

New isolated nucleic acid derived from Enterococcus faecium encoding an Enterococcus faecium polypeptide useful for detection, prevention and treatment of a pathological condition resulting from a bacterial infection.

Example 1; SEQ ID NO 6487; 243pp; English.

The invention relates to an isolated nucleic acid derived from Enterococcus faecium encoding an Enterococcus faecium polypeptide having one of 10 fully defined sequences given in the (or comprising 40 sequential nucleotides chosen from any of the nucleic acids, its complement or sequences hybridising to it). Also included are a recombinant vector comprising the nucleic acid operably linked to a transcription regulatory element, a cell comprising the vector and a single-stranded probe comprising the nucleic acid. The nucleic acids are chosen from 3654 disclosed sequences encoding 3654 disclosed proteins. The nucleic acid is useful for diagnosing pathological conditions resulting from E. faecium bacterial infection (e.g. urinary tract infection, bacteraemia, endocarditis, wounds and abdominal-pelvic infection) and for screening drugs such as agonists and antagonists. The

CC nucleic acid is useful for recombinant production of Candida albicans -  
CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
CC and vaccines containing the nucleic acid are useful for preventing or  
CC treating Enterococcus faecium infections. The present sequence represents  
CC one if the disclosed E. faecium proteins.

XX SQ Sequence 156 AA;

Query Match 10.8%; Score 91.5; DB 7; Length 156;  
Best Local Similarity 23.0%; Pred. No. 0.2;  
Matches 35; Conservative 28; Mismatches 74; Indels 15; Gaps 5;  
6 IYFAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLDKWA 65  
14 IYLAGPFSEQIDRVSRIEKALENKTIV--TSFYSPRHQ-----ESNVELFSAGWA 64  
66 TATYNNDLINGIKTNDIMLVGVIPDEE--DVGLGMELGVALSQKYVLLVLPDEDYKGPIN 123  
65 QEVYEKMEELTNAEFVVAILDPEHQTIDPGTAYELGVATMLKKPMIIV---QEETVPTN 121  
124 LMSWGVSDNVIKMSQ-LKDFNFKPRDFYEG 154  
122 LMITQSLHTLYLKSDQAVREYDFETLPVETVYG 153

RESULT 14

ADW47705  
ID ADW47705 standard; protein; 719 AA.

XX AC ADW47705;

XX DT 24-MAR-2005 (first entry)

XX DE F. magna protein L.

XX KW immunogenicity; immune disorder; cystic fibrosis; infection; cancer;  
XX KW neoplasm; respiratory disorder; antibacterial; cytostatic;  
XX KW immunosuppressive; cns-gen.; respiratory-gen.; immunoglobulin.

XX OS Finegoldia magna.

XX PN WO2005001480-A1.

XX PD 06-JAN-2005.

XX PF 28-JUN-2004; 2004WO-AU000856.

XX PR 27-JUN-2003; 2003AU-00903317.

XX PA (PROT-) PROTEOME SYSTEMS INTELLECTUAL PROPERTY P.

XX PI Pedersen SK, Cole RA, Weinberger R, Sloane AJ;

XX WPI; 2005-101297/11.

XX Identifying immunogenic protein useful as antigen in vaccine composition,  
PT involves obtaining protein complex comprising immunoglobulin or  
PT immunoglobulin-containing fraction from subject and identifying protein  
PT bound to immunoglobulin.

XX Disclosure; SEQ ID NO 3; 158pp; English.

XX The invention relates to a method of identifying an immunogenic protein  
CC (immunoglobulin) or its fragment capable of eliciting an immune response.  
CC The method is useful for identifying an immunogenic protein or its  
CC fragment capable of eliciting an immune response in a subject. The method  
CC is useful for treating an autoimmune condition (e.g. rheumatoid arthritis  
CC or multiple sclerosis), cystic fibrosis, acute pulmonary exacerbation,  
CC infection (e.g. P. aeruginosa) and cancer. The method is useful for  
CC identifying a marker of a condition or in diagnosis of a condition such  
CC as disease or disorder e.g., infectious disease, cancer, or autoimmune  
CC disease. In the method, the protein or peptide fragment are sufficiently  
CC non-degraded to permit determination of their amino acid sequences. The

CC present sequence represents the amino acid sequence of the F. magna  
CC antibody light chain binding protein, protein L.  
SQ Sequence 719 AA;

Query Match 10.7%; Score 91; DB 9; Length 719;  
Best Local Similarity 28.0%; Pred. No. 2.1;  
Matches 37; Conservative 17; Mismatches 50; Indels 28; Gaps 6;

Qy 11 GWFTDRQNKAYKEAMEALKENP--TIDLENSYVPLDNQYKGI RVDHEPHE----- 57  
Db 349 GTFPAEATAEAYRYADLLAKENGKYTADLEDGGYTINIRFAGKVDKPEKEQVTIKENI 408  
Qy 58 YLHDKVWATATYNNDL-----NGIKTNDIML---GVYIPDEEDVGLGMEIGYALSQKYYV 109  
Db 409 YFEDGTQVATATFKGTFPAEATAEAYRYADLLSKHGKYTADLEDGGYTINIRFA---GKEE 465  
Qy 110 LLVIPDEDYGKP 121  
Db 466 -----PEETPEKP 473

RESULT 15  
ABB54050  
ID ABB54050 standard; protein; 462 AA.  
AC ABB54050;  
XX  
XX 29-AUG-2003 (revised)  
DT 16-MAY-2002 (first entry)  
XX  
XX Lactococcus lactis protein yheD.  
DE  
XX  
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
XX  
XX Lactococcus lactis; IL1403.  
XX  
XX FR2807446-A1.  
XX  
XX 12-OCT-2001.  
XX  
XX 11-APR-2000; 2000FR-00004630.  
XX  
XX 11-APR-2000; 2000FR-00004630.  
XX  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX  
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
XX  
XX WPI; 2002-043418/06.  
XX  
XX New nucleotide sequence useful in the identification or Lactococcus  
XX lactis and related species.  
XX  
XX Claim 6; SEQ ID NO 752; 2504pp; French.  
XX

The present invention is related to a Lactococcus lactis nucleotide  
sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic  
acid sequence is useful in the detection and/or amplification of nucleic  
acid sequence, particularly to identify Lactococcus lactis or related  
species. The proteins of the invention are useful for the biosynthesis or  
biodegradation of a composition of interest. The invention helps research  
in lactic bacteria, particularly useful in the production of yogurt and  
cheese. Note: The sequence data for this patent is based on equivalent  
patent WO200177334 (published 18-OCT-2001) which is available in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 29-AUG-2003 to  
CC standardise OS field)  
XX  
XX Sequence 462 AA;

Query Match 10.2%; Score 86.5; DB 5; Length 462;  
Best Local Similarity 24.1%; Pred. No. 3.4;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:56:48 ; Search time 47 Seconds  
(without alignments)  
276.172 Million cell updates/sec

Title: US-10-049-750-14

Perfect score: 848

Sequence: 1 MPKTYIFGAGWFTDRQNK.....QLKDFNFKPRDFYEGAVY 157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.\*
- 1: /cgn2\_6/ptodata/1/1aa/5 COMB.pcp.\*
  - 2: /cgn2\_6/ptodata/1/1aa/6 COMB.pcp.\*
  - 3: /cgn2\_6/ptodata/1/1aa/H COMB.pcp.\*
  - 4: /cgn2\_6/ptodata/1/1aa/pCTUS COMB.pcp.\*
  - 5: /cgn2\_6/ptodata/1/1aa/RE COMB.pcp.\*
  - 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.5	10.8	156	2	US-09-107-532A-6487
2	84.5	10.0	396	2	US-09-800-729-86
3	83.5	9.8	382	1	US-08-477-451-28
4	83.5	9.8	1781	1	US-08-477-451-11
5	82.5	9.7	766	4	PCT-US94-00198-6
6	82.5	9.7	920	2	US-09-248-796A-17529
7	80.5	9.5	2337	1	US-08-354-973-1
8	78.5	9.3	170	2	US-09-134-000C-4783
9	76.5	9.0	338	2	US-09-489-039A-11745
10	76	9.0	279	2	US-09-134-001C-4707
11	76	9.0	434	1	US-08-795-475-3
12	76	9.0	434	1	US-08-325-278B-3
13	75.5	8.9	181	1	US-08-307-499-58
14	75.5	8.9	181	2	US-09-299-268-58
15	75.5	8.9	537	2	US-09-248-796A-19446
16	75	8.8	160	2	US-09-107-532A-4152
17	74	8.7	171	2	US-10-085-572-7
18	74	8.7	171	2	US-10-087-195-7
19	74	8.7	190	2	US-09-618-919A-18
20	74	8.7	750	2	US-09-248-796A-18005
21	74	8.7	993	2	US-08-836-687B-30
22	74	8.7	1646	2	US-09-535-008-67
23	74	8.7	1647	2	US-09-535-008-2
24	74	8.7	1647	2	US-09-824-574-4
25	74	8.7	1647	2	US-09-538-092-1172
26	74	8.7	1649	2	US-09-535-008-75
27	74	8.7	1650	2	US-09-535-008-71

Sequence 9752, Ap  
Sequence 69, Appl  
Sequence 65, Appl  
Sequence 77, Appl  
Sequence 73, Appl  
Sequence 1948, Ap  
Sequence 4434, Ap  
Sequence 6783, Ap  
Sequence 47, Appl  
Sequence 45, Appl  
Sequence 28646, A  
Sequence 7334, Ap  
Sequence 27095, A  
Sequence 4391, Ap  
Sequence 192, App  
Sequence 11595, A  
Sequence 99, Appl

28 74 8.7 1659 2 US-09-949-016-9752  
29 74 8.7 1678 2 US-09-535-008-69  
30 74 8.7 1679 2 US-09-535-008-85  
31 74 8.7 1681 2 US-09-535-008-77  
32 74 8.7 1682 2 US-09-535-008-73  
33 73.5 8.7 252 2 US-09-710-279-1948  
34 73.5 8.7 308 2 US-09-134-001C-4434  
35 73.5 8.7 499 2 US-09-328-352-6783  
36 73.5 8.7 844 2 US-09-422-936-47  
37 73.5 8.7 960 2 US-09-422-936-45  
38 73 8.6 225 2 US-09-252-991A-28646  
39 73 8.6 264 2 US-09-489-039A-7334  
40 73 8.6 992 2 US-09-252-991A-27095  
41 72.5 8.5 164 2 US-09-107-532A-4391  
42 72.5 8.5 331 2 US-09-370-838-192  
43 72.5 8.5 331 2 US-09-854-133-192  
44 72.5 8.5 631 2 US-09-949-016-11595  
45 72.5 8.5 659 2 US-10-152-886-99

ALIGNMENTS

RESULT 1  
US-09-107-532A-6487  
; Sequence 6487, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arinello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 6487:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...156  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6487:  
US-09-107-532A-6487

Query Match 10.8%; Score 91.5; DB 2; Length 156;  
Best Local Similarity 23.0%; Pred. No. 0.021;  
Matches 35; Conservative 28; Mismatches 74; Indels 15; Gaps 5;  
  
QY 6 IYFAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGI RVD EHP EY LHD K V A 65  
DB 14 IYLAGPFFSEQIDRVSRIKALEENKTIV--TSFYSPRHQ-----ESNYLFSAGWA 64  
  
QY 66 TATYNNDLNGIKTNDIMLGVIYIPDEE--DVGLGMEGYALSQGYVLLVIPDEDYDKPIN 123  
DB 65 QEYVEKDMBELTNAEFVVAILDHFHQIDPGTAVELGVATMLKKPMIIV---QEETVPTN 121  
  
QY 124 LMSWGVSDNVKMSQ-LKDFNFKPRDFYEG 154  
DB 122 LMITQSLHTLYKSDQAVREYDFETLPVETVG 153  
  
RESULT 2  
US-09-800-729-86  
; Sequence 86, Application US/09800729  
; Patent No. 6605592  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: PZ044P1  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: 60/155,709  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 86  
; LENGTH: 396  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-800-729-86

Query Match 10.0%; Score 84.5; DB 2; Length 396;  
Best Local Similarity 23.1%; Pred. No. 0.52;  
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;  
  
QY 9 GAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHP EY LHD K V 63  
DB 251 GDRWILTAATITYPKDSVSLRKQSVNVFLGHTAIDEMLKGHPVHRVHVVDYRQNE- 309  
  
QY 64 WATATYNNDL-----NGIKTNDIMLGVIYIPDEDV-----GLGMEGYALSQGYK 107  
DB 310 --SHNFGSDIALLELQHSILPGPNVLPCLPDNETLYRSGLLGYVSGFGMEMGLTTELK 367  
  
QY 108 YVLL-VIPDE 116  
DB 368 YSRLPVAPRE 377

RESULT 3  
US-08-477-451-28  
; Sequence 28, Application US/08477451  
; Patent No. 5928865  
; GENERAL INFORMATION:  
; APPLICANT: Covacci, Antonello  
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,451  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: McClung, Barbara G.  
REGISTRATION NUMBER: 33,113  
REFERENCE/DOCKET NUMBER: 0335.002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-601-2708  
TELEFAX: 510-655-3542  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 382 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-477-451-28  
  
Query Match 9.8%; Score 83.5; DB 1; Length 382;  
Best Local Similarity 26.0%; Pred. No. 0.64;  
Matches 38; Conservative 22; Mismatches 71; Indels 15; Gaps 5;  
  
QY 17 ONKAYKEAMEALKENPTID--LENSYVPLDNOYKGI RVD EHP EY LHD K V A T A T Y N N D L 73  
DB 111 KSKAYGDLSQLADNPLKYKLLPNFYAVLUNQTTFTKEKERLSYYPQVKTSSEIKTT 170  
  
QY 74 NGIKTND---IMLGVIYIPDEDVGLGMEGYALSQGYVLLVIPDEDYDKPI-NLMSWGV 129  
DB 171 ATTKRAQALLQMGVFLDDEEQNKASRLALSQKA-----IBEYSNNISNLLSRKE 222  
  
QY 130 SDNVKMSQLKDFNFKPRDFYEGA 155  
DB 223 LDNIYYLQLERNKFDKAKDIAQA 248

RESULT 4  
US-08-477-451-11  
; Sequence 11, Application US/08477451  
; Patent No. 5928865  
; GENERAL INFORMATION:  
; APPLICANT: Covacci, Antonello  
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,451  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: McClung, Barbara G.  
REGISTRATION NUMBER: 33,113  
REFERENCE/DOCKET NUMBER: 0335.002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-601-2708  
TELEFAX: 510-655-3542  
INFORMATION FOR SEQ ID NO: 11:







Qy	6	IYFAG----	WTDRQNKAYKEAMEALAKENFTIDLENSYVPLDNQYKGIRVDEHPVELHDK	62
Db	15	IPFASGAMTFTIDKNK-----	YDNRHYTKTINSKIEHL--S	49
Qy	63	VWATATYNDELNGIKTNDIMLGVIYPBEDVGL-----	GMEGLVAL-----	103
Db	50	ISTVTNTNIIISGGK-----	LAVFTTGDKINVTYNNKRLSIKEKRAVDORGVLGNPNPH	104
Qy	104	SQCKYVLLVIPDEDYKGPINLMSW-----	GVS---DNVI--KMSQLKDFFNFNK	146
Db	105	NNRKLTIIVPEKDL-KSLNIQSLLGEIDLNQVNKLKHVSLETDRITQLKRESELNQVNIES		163
Qy	147	PRDFPY	152	
Db	164	SKANFY	169	

RESULT 11  
 US-08-795-475-3  
 ; Sequence 3, Application US/08795475  
 ; Patent No. 5965390  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bjvrck, Lars  
 ; APPLICANT: Sjvbring, Ulf  
 ; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: SEED AND BERRY LLP  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: USA  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/795,475  
 ; FILING DATE: 11-FEB-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mcmasters, David D.  
 ; REGISTRATION NUMBER: 33,963  
 ; REFERENCE/DOCKET NUMBER: 100084.402D1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 434 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Escherichia coli LE392/pHDLG, DSM 7055  
 ; US-08-795-475-3

```

RESULT 12
US-08-325-278B-3
; Sequence 3, Application US/08325278B
; Patent No. 6822075
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed IP Law Group
; STREET: 701 Fifth Avenue Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,278B
; FILING DATE: 26-Oct-1994
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 100084.402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli LE392/pHDLG, DSM 7055
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-08-325-278B-3

Query Match          9.0%; Score 76; DB 2; Length 434;
Best Local Similarity 35.8%; Pred. No. 5.5;
Matches 24; Conservative 9; Mismatches 26; Indels 8; Gaps 2;

QY 11 GWFTDRQNKYAKMEALKENP--TTDLNSVYPLDNOYKGRVDSHPHYLHDKKWATAT 68
DB 256 GFPAETARAYADLLAKENGYTADLEGGYTIINRFAGKKVKPEEPM-----T 309
QY 69 YNNDLNG 75
DB 310 YKLLNG 316

RESULT 13
US-08-307-499-58
; Sequence 58, Application US/08307499
; Patent No. 5651972
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik

```

```
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,499
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-307-499-58

Query Match      8.9%; Score 75.5; DB 1; Length 181;
Best Local Similarity 25.8%; Pred. No. 1.7;
Matches 32; Conservative 24; Mismatches 31; Indels 37; Gaps 8;

Qy 44 DNOYKGRVDEHPHYLHDKVMATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMELG--- 100
   |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 47 DNRYGNDAV-----YTHDKCVISAVSTDSLFDIK--DTL-----DDVDI-VGIDEGQFF 92

Qy 101 -----YALSOGKYVLLVIPDEDY-----GKPINLMSWGVSDNVIKMSQL-----KDF 142
   |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 93 NDIVEFCEYIANKGKIVIVAALDGTVERKPFNGNLNLP--LSEKVTKLNAICMICH RDA 150

Qy 143 NFNK 146
   :|  :|
Db 151 SFSK 154

RESULT 14
US-09-299-268-58
; Sequence 58, Application US/09299268
; Patent No. 6217882
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E. P. J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
```

```
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,268
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/901,127
; FILING DATE:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-299-268-58

Query Match      8.9%; Score 75.5; DB 2; Length 181;
Best Local Similarity 25.8%; Pred. No. 1.7;
Matches 32; Conservative 24; Mismatches 31; Indels 37; Gaps 8;

Qy 44 DNOYKGRVDEHPHYLHDKVMATATYNNDLNGIKTNDIMLGVIYPDEEDVGLGMELG--- 100
   |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 47 DNRYGNDAV-----YTHDKCVISAVSTDSLFDIK--DTL-----DDVDI-VGIDEGQFF 92

Qy 101 -----YALSOGKYVLLVIPDEDY-----GKPINLMSWGVSDNVIKMSQL-----KDF 142
   |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
Db 93 NDIVEFCEYIANKGKIVIVAALDGTVERKPFNGNLNLP--LSEKVTKLNAICMICH RDA 150

Qy 143 NFNK 146
   :|  :|
Db 151 SFSK 154

RESULT 15
US-09-248-796A-19446
; Sequence 19446, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19446
```

```
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19446

Query Match      8.9% Score 75.5; DB 2; Length 537;
Best Local Similarity 26.5%; Pred. No. 8.7;
Matches 41; Conservative 18; Mismatches 59; Indels 37; Gaps 10;

Qy  2 PKKTIYFGAGW-FTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGIRVDEHPEYLH 60
Db  314 PKPTDEF--LWEFIDKCGK-YK---SPKTNP-LDFYKTYVRFD-----ASEHFSLIH 358

Qy  61 D-----KWTATATYNDLNGIKTNDIMLGV-----YIPDEEDVGLGMELGYALSQ 105
Db  359 DPRNEYNKLTYTVERLNNIFGGKPIEYINLEIDEIKQVAIKMLKONEPVFFGSDVG-KFSD 417

Qy  106 GKYYLLVIPDEDDYGKPINLMSWGSVDNVIKMSQLK 140
Db  418 SKSGILDTTAYDYS-----TAFDFSLDITKLQRLK 447
```

Search completed: December 23, 2005, 22:05:47  
Job time : 50 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 22:04:24 ; Search time 165 Seconds  
(without alignments)  
397.571 Million cell updates/sec

Title: US-10-049-750-14  
Perfect score: 848  
Sequence: 1 MPKXITFGAGWFTDRQNK.....QLKDFNFKRPFDFEGAVY 157

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:  
1: /cgm2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgm2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgm2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgm2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgm2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgm2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	10.1	464	3	US-09-808-602-95
2	85.5	10.1	487	3	US-09-808-602-93
3	85.5	10.1	487	3	US-09-800-198-80
4	84.5	10.0	396	3	US-09-800-729-86
5	84.5	10.0	396	3	US-09-833-245-2176
6	84.5	10.0	452	3	US-09-808-602-21
7	84.5	10.0	487	3	US-09-808-602-17
8	84.5	10.0	487	3	US-09-808-602-19
9	84.5	10.0	487	3	US-09-800-198-17
10	84.5	10.0	487	3	US-09-800-198-19
11	84	9.9	1537	6	US-11-097-143-27759
12	84	9.9	1634	5	US-10-732-923-8723
13	84	9.9	1638	5	US-10-732-923-8724
14	84	9.9	1638	6	US-11-097-143-12630
15	83	9.8	3354	4	US-10-160-758-11
16	83	9.8	3354	4	US-10-160-758-12
17	83	9.8	3354	4	US-10-174-677-9
18	83	9.8	3354	4	US-10-080-334-197
19	83	9.8	3354	4	US-10-080-334-198
20	83	9.8	3399	4	US-10-080-334-196
21	82.5	9.7	376	4	US-10-335-977-8864
22	82.5	9.7	376	4	US-10-335-977-8865
23	82.5	9.7	382	4	US-10-335-977-8866
24	82.5	9.7	420	4	US-10-437-963-180891
25	82	9.7	1638	5	US-10-732-923-8722
26	81	9.6	241	5	US-10-472-928-1270
27	81	9.6	783	4	US-10-282-122A-48649

28	80.5	9.5	2233	4	US-10-369-493-1998	Sequence 1998, Ap
29	79.5	9.4	367	4	US-10-425-115-294967	Sequence 294967, A
30	79.5	9.4	430	4	US-10-425-114-38846	Sequence 38846, A
31	79.5	9.4	934	5	US-10-840-512-228	Sequence 228, App
32	79.5	9.4	1452	4	US-10-282-122A-51753	Sequence 51753, A
33	79.5	9.4	1923	5	US-10-450-763-55533	Sequence 55533, A
34	79	9.3	662	5	US-10-732-923-8221	Sequence 8221, Ap
35	78	9.2	333	4	US-10-435-226-2	Sequence 2, Appli
36	78	9.2	628	4	US-10-389-566-1068	Sequence 1068, Ap
37	78	9.2	628	5	US-10-732-923-9638	Sequence 9638, Ap
38	78	9.2	1078	5	US-10-505-818-15	Sequence 15, Appl
39	77.5	9.1	206	5	US-10-732-923-8422	Sequence 8422, Ap
40	77	9.1	402	5	US-10-732-923-10596	Sequence 10596, A
41	77	9.1	402	5	US-10-732-923-10617	Sequence 10617, A
42	76.5	9.0	288	4	US-10-425-115-248274	Sequence 248274, A
43	76.5	9.0	406	5	US-10-501-282-1070	Sequence 1070, Ap
44	76.5	9.0	473	4	US-10-424-599-227005	Sequence 227005, A
45	76.5	9.0	594	4	US-10-425-115-248278	Sequence 248278, A

ALIGNMENTS

RESULT 1  
US-09-808-602-95  
; Sequence 95, Application US/09808602  
; Patent No. US20020155115A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Corine A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Herrman, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishnu  
; APPLICANT: Mezes, Peter S  
; APPLICANT: MacDougall, John  
; TITLE OF INVENTION: No. US20020155115A1e1 Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-697 CIP  
; CURRENT APPLICATION NUMBER: US/09/808,602  
; CURRENT FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 95  
; LENGTH: 464  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-808-602-95

Query Match	10.1%	Score 85.5;	DB 3;	Length 464;
Best Local Similarity	23.1%	Pred. No. 3.9;	Mismatches 49;	Indels 25;
Matches	30;	Conservative 26;	Gap 5;	
QY	9	GAGFTDQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPEYLDHKV	63	
Db	251	GDRMLTAAHTVYPKDSVSLKQSVNVLGHTAIDMLKGNHPVHRVWVHPDYRQNE-	309	
QY	64	WATATYNNDL-----NGIKTNDIMLGVIYIDDEEV-----GLGMELGYALSQOK	107	
Db	310	--SHNFSGDIALLELQHSIPLGNVLPVCLPDNETLYRSGLLYGVSGFGMENGMLTTELK	367	
QY	108	YVLL-VIPDE 116		
Db	368	YSRLPVAPRE 377		

RESULT 2  
US-09-808-602-93  
; Sequence 93, Application US/09808602  
; Patent No. US20020155115A1

```
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Heriman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 93
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-93

Query Match      10.1%; Score 85.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 4.2;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

QY      9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSVYVPLDNOYK-----GIRVDEHPEYLDKVK 63
DB      274 GDRWILTAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKGNHPVHRVVVHDPYRQNE- 332
QY      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQKG 107
DB      333 --SHNFGDIALLELQHSIPLGPNVLPVCLPDNETLYRSLGLGYVSGFGMGWLTTELK 390
QY      108 YVLL-VIPDE 116
DB      391 YSRLPVAPRE 400

RESULT 4
US-09-800-729-86
; Sequence 86, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 86
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-86

Query Match      10.0%; Score 84.5; DB 3; Length 396;
Best Local Similarity 23.1%; Pred. No. 4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

QY      9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSVYVPLDNOYK-----GIRVDEHPEYLDKVK 63
DB      251 GDRWILTAHTVYPKDSVSLRKNQSVNVFLGHTAIDEMLKGNHPVHRVVVHDPYRQNE- 309
QY      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQKG 107
DB      310 --SHNFGDIALLELQHSIPLGPNVLPVCLPDNETLYRSLGLGYVSGFGMGWLTTELK 367
QY      108 YVLL-VIPDE 116
DB      368 YSRLPVAPRE 377

RESULT 5
US-09-833-245-2176
; Sequence 2176, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2176
; LENGTH: 396
; TYPE: PRT

; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herimann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-80

Query Match      10.1%; Score 85.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 4.2;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;
```

```
; ORGANISM: Homo sapiens
US-09-833-245-2176

Query Match      10.0%; Score 84.5; DB 3; Length 396;
Best Local Similarity 23.1%; Pred. No. 4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPHYLHDKV 63
Db 251 GDRWILTAHTIYPKDSVSLRKQSNVFLGHTAIDMLKLGHPVHRVWVHPDYRQNE- 309
Qy 64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db 310 --SHNFGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMGWLTTELK 367
Qy 108 YVLL-VIPDE 116
Db 368 YSRLPVAPRE 377

RESULT 6
US-09-808-602-21
; Sequence 21, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; PRIOR FILING DATE: 2001-03-14
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-21

Query Match      10.0%; Score 84.5; DB 3; Length 452;
Best Local Similarity 23.1%; Pred. No. 4.8;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPHYLHDKV 63
Db 239 GDRWILTAHTIYPKDSVSLRKQSNVFLGHTAIDMLKLGHPVHRVWVHPDYRQNE- 297
Qy 64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db 298 --SHNFGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMGWLTTELK 355
Qy 108 YVLL-VIPDE 116
Db 356 YSRLPVAPRE 365

RESULT 7
US-09-808-602-17
; Sequence 17, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
```

```
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR FILING DATE: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-17

Query Match      10.0%; Score 84.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 5.4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPHYLHDKV 63
Db 274 GDRWILTAHTIYPKDSVSLRKQSNVFLGHTAIDMLKLGHPVHRVWVHPDYRQNE- 332
Qy 64 WATATYNNDL-----NGIKTNDIMLGVIYPDEEDV-----GLGMELGYALSQK 107
Db 333 --SHNFGDIALLELQHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMGWLTTELK 390
Qy 108 YVLL-VIPDE 116
Db 391 YSRLPVAPRE 400

RESULT 8
US-09-808-602-19
; Sequence 19, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR FILING DATE: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 2001-03-05
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-19

Query Match      10.0%; Score 84.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 5.4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy 9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPHYLHDKV 63
```

```
Db      274 GDRWILTAHTIYPKDSVSRKQSVNVFLGHTAIDEMKLGNHPVHRVVHDPYRQNE- 332
Qy      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEBDV-----GLGMELGYALSQGK 107
Db      333 --SHNFGSDIALLELQHSIPLGPNVLPVCLPDNETLYRSLGLLVGVSFGMGWLTTELK 390
Qy      108 YVLL-VIPDE 116
Db      391 YSRLPVAPRE 400

RESULT 9
US-09-800-198-17
; Sequence 17, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Cornie AM
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 17
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-17
```

```
Query Match      10.0%; Score 84.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 5.4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy      9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPEYLHDKV 63
Db      274 GDRWILTAHTIYPKDSVSRKQSVNVFLGHTAIDEMKLGNHPVHRVVHDPYRQNE- 332
Qy      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEBDV-----GLGMELGYALSQGK 107
Db      333 --SHNFGSDIALLELQHSIPLGPNVLPVCLPDNETLYRSLGLLVGVSFGMGWLTTELK 390
Qy      108 YVLL-VIPDE 116
Db      391 YSRLPVAPRE 400
```

```
RESULT 10
US-09-800-198-19
; Sequence 19, Application US/09800198
; Publication No. US20030087816A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Cornie AM
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrmann, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishna
; APPLICANT: Mezes, Peter S
; APPLICANT: Rastelli, Luca
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-697
; CURRENT APPLICATION NUMBER: US/09/800,198
; CURRENT FILING DATE: 2001-03-05
```

```
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 19
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-198-19
```

```
Query Match      10.0%; Score 84.5; DB 3; Length 487;
Best Local Similarity 23.1%; Pred. No. 5.4;
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

Qy      9 GAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYK-----GIRVDEHPEYLHDKV 63
Db      274 GDRWILTAHTIYPKDSVSRKQSVNVFLGHTAIDEMKLGNHPVHRVVHDPYRQNE- 332
Qy      64 WATATYNNDL-----NGIKTNDIMLGVIYPDEBDV-----GLGMELGYALSQGK 107
Db      333 --SHNFGSDIALLELQHSIPLGPNVLPVCLPDNETLYRSLGLLVGVSFGMGWLTTELK 390
Qy      108 YVLL-VIPDE 116
Db      391 YSRLPVAPRE 400
```

```
RESULT 11
US-11-097-143-27759
; Sequence 27759, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27759
; LENGTH: 1537
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-27759
```

```
Query Match      9.9%; Score 84; DB 6; Length 1537;
Best Local Similarity 20.4%; Pred. No. 30;
Matches 42; Conservative 30; Mismatches 48; Indels 86; Gaps 11;

Qy      9 GAGWFTDRQ-----NKAYKEAMEALKENPTI-----DLENSYVPLDN 45
Db      584 GWDWIDDEEDSCGSDNDHDKPKVE---EQPTATEDATDKAQTGNDDEKDLITKAKVEDD 640
Qy      46 QYKGIRVDEHPEY-----LHDKV-----WATATYNNDLNGIKTN 79
```



[illegible]

## RESULT 12

```

US-10-732-923-8723
; Sequence 8723, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8723
; LENGTH: 1634
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-732-923-8723

```

Query Match 9.9%; Score 84; DB 5; Length 1634;

Query	Best Local	Similarity	Pred.
Query	Best Local	Similarity	Pred.

Best local similarity	20.44; Acc: 32%	Matches	42; Conservative	30; Mismatches	48; Indels	86; Gaps	11; Gaps

Qy	9	GAGWFTDRQ-----NKAYKEAMEALKENPTI-----	-----DLENSYVPLDN 45
Db	681	GWMDIDDESDSCGSDNDHKKPYE---EQFTATEDATDKAQTGDNDEDAKLITKAKVEDD 737	
Qy	46	QYKGIKRVDSHPEY-----LHDKV-----	-----WATATVNDNLNGIKTN 79
Db	738	EY---RTESQTYTSTAHTIHEKVQASIMVNGTLKEYQIKGLEWLSVLTNNNLINGILAD 794	
Qy	80	DIMLG-----VYIPDEEDVGLGMELGVALSQGKYVLLAV-----	IPD-----EDYKGP 121
Db	795	EMGLGKTIQTSILVTLMDRKKV-----	-----MGPYLIIVPLSTLPNVLFSEFKWAPA 843
Qy	122	INLMSWGVSDNVTKM---SOLKDFNFN 145	
Db	844	VGVVSYKGGSPQGGRRLLQNMRATKFN 869	

RESULT 13

```

US-10-732-923-8724
; Publication 8724, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS
; FILING DATE: 2003-12-10
; CURRENT APPLICATION NUMBER: US/10/732
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8724
; LENGTH: 1638
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-732-923-8724

```

Query Match	9.9%	Score 84	DB 5	Lenath 1638
-------------	------	----------	------	-------------

Query Match	5.5%;	score
Best Local Similarity	20.4%;	pred.

	Matches	42; Conservative	30; Mismatches	48; Indels	86; Gaps	11;
Qy	9	GAGWFFTRQ----	NKAYKEAMEALKENPTI-----	DLENSYVPLDN	45	
		: : : : : :	: : : : : :			
Db	685	GWDWIDDEEDSCGSDNDHKPKVE--	EQTATEDATDKAQTGNDEDAKOLITKAKVEDD	.741		
		: : : : : :	: : : : : :			
Qy	46	OYKGI RVDEHPEY-----	LHDKV-----	WATATYNDLNGIKTN	79	
		: : : : : :	: : : : : :			
Db	742	EY---RTBEQTYYSIAHTHEKRYVEQASIMVNGTLKEYQIKGLEWLSLYNNLNNGLIAD	798			
		: : : : : :	: : : : : :			
Qy	80	DIMLG-----	VPIPEEDYGLGELGYALSQKGYVLLV---	IPD-----	EDYGKP	121
		: : : : : :	: : : : : :			
Db	799	EMGLGKTIQITISLVITMDRKKV-----	MGPLYIIIVPLSTLPNHWLEFEKWA	847		
		: : : : : :	: : : : : :			
Qy	122	INLMSGVGSDNVIKMY--	SQDKDFNFN	145		
		: : : : : :	: : : : : :			
Db	848	VGVSYSYKGSPOGRBLLONOMRAKFN	873			
		: : : : : :	: : : : : :			

## RESULT 14

```

US-11-097-143-12630
; Sequence 12630, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH A
; TITLE OF INVENTION: ARRAYS, FOR DETECTION
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12630
; LENGTH: 1638
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-12630

```

Query Match 9.9%; Score 84; DB 6; Length 1638;

Query Match	5.50;	Score 31;
Best Local Similarity	20.48;	Pred. No. 32;

Best local similarity 40.0, 30; Mismatches 48; Indels 86; Gaps 11	
Matches 42; Conservative	
Qy	9 GAGFTDRQ-----NKAYKEAMEALKENPTI-----DIENSVYPLDN 45
Db	685 GWDWIDDEEDSCGSDDDHKPVE--EPTATEDATDKAQTGDNDEDAKLITRAKVEDD 741
Qy	46 QYKQIRVDEHPY-----LHDKV-----WATATYNDLNGIKTN 79
Db	742 EY---RTEEQTYYSIAHTIHEKVVEQASIMVNGTLKEYQIKGLEWLSVLYNNLNGIAD 798
Qy	80 DIMLG-----VYIPDEEDVGLGMELGYALSGQKYVLLV---IPD-----EDYCKP 121
Db	799 EMGLGKTIQTISLVTYIMDRKKV-----MGPYLIIVPLSTLPNWWLFEFKWAPA 847
Qy	122 INLMSWGVSDNVIKW---SQLKDFNFN 145

```
Db      848  VGVSYKGSFGRRLLQNMRA TKFN 873
          : : : | : : : | : : : ||
RESULT 15
US-10-160-758-11
; Sequence 11, Application US/10160758
; Publication No. US20030036076A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-089C
; CURRENT APPLICATION NUMBER: US/10/160,758
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 3354
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-160-758-11

Query Match      9.8%; Score 83; DB 4; Length 3354;
Best Local Similarity 22.2%; Pred.No. 1.1e+02;
Matches 36; Conservative 32; Mismatches 68; Indels 26; Gaps 8;

Qy      15  DRQNKAYKEAMEALKE--NPTIDL-----ENSYVPLDQ-QYKIRVDEHPEYLHD 61
          || : : : : : : : || : : : : : : : || : : : : : : : || : : : : : : :
Db      1593 DRERQSFYHLVAIVEDSGTPTLSATTHVYTVIVDENDNAPMFQQPHYEVLLDEGPDTLNT 1652

Qy      62  KWTATATYND--LNGIKTWNDIMLGVIYPD---EEDVGL---GMELGYALSQGYVLLVI 113
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1653 SLITIQALDLDEGPNGTVTVAIVAGNIVNTFRIDRHMGVITAAKELDYEISHGRYTLIVT 1712

Qy      114  PDEDYGKPINLMSGVSDNVIKMSQLKDFNFKPRD-FYEG 154
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1713 ATDQ-----CPILSHRLTSTTTVLNVNNDINDNVPTFPDRDYEG 1750

Search completed: December 23, 2005, 22:20:26
Job time : 167 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 22:05:00 ; Search time 12 Seconds  
(without alignments)  
93.321 Million cell updates/sec

Title: US-10-049-750-14  
Perfect score: 848  
Sequence: 1 MPKTIYFGAGWFTDRQNKA.....QLKDFNFKPRDFYEGAVY 157

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 54001

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	78	9.2	750	7	US-11-070-627-4
2	73.5	8.7	252	6	US-10-793-626-1948
3	73	8.6	756	7	US-11-074-176-202
4	70.5	8.3	750	7	US-11-070-627-3
5	69.5	8.2	391	6	US-10-979-821-12
6	69.5	8.2	750	7	US-11-070-627-2
7	69.5	8.2	786	7	US-11-070-627-9
8	68.5	8.1	395	6	US-10-793-626-668
9	68.5	8.1	1107	6	US-10-485-517-145
10	68	8.0	605	7	US-11-094-586-4
11	68	8.0	747	6	US-10-821-234-1662
12	68	8.0	750	7	US-11-070-627-1
13	68	8.0	751	6	US-10-821-234-1007
14	67.5	8.0	1302	7	US-11-004-057-6
15	67.5	8.0	1493	6	US-11-004-057-21
16	67	7.9	485	6	US-10-630-203-2
17	67	7.9	485	7	US-11-103-037-1
18	67	7.9	1048	6	US-10-392-234A-14
19	66.5	7.8	959	6	US-10-467-962B-4
20	66.5	7.8	1170	6	US-10-858-730-71
21	66.5	7.8	1316	7	US-11-091-643-4
22	66	7.8	770	7	US-11-070-627-8
23	66	7.8	770	7	US-11-070-627-10
24	66	7.8	932	7	US-11-017-550-65
25	65.5	7.7	385	6	US-10-873-528-144

Sequence 8376, Ap  
Sequence 1536, Ap  
Sequence 20, Appl  
Sequence 5510, Ap  
Sequence 604, App  
Sequence 24, Appl  
Sequence 26, Appl  
Sequence 22, Appl  
Sequence 1004, Ap  
Sequence 20, Appl  
Sequence 256, App  
Sequence 2, Appl  
Sequence 1532, Ap  
Sequence 18, Appl  
Sequence 4, Appl  
Sequence 358, App  
Sequence 101, App  
Sequence 3, Appl  
Sequence 2368, Ap

## ALIGNMENTS

## RESULT 1

US-11-070-627-4  
; Sequence 4, Application US/11070627  
; Publication No. US20050271625A1  
; GENERAL INFORMATION:  
; APPLICANT: Nash, Kevin R.  
; APPLICANT: Burger, Corinna  
; TITLE OF INVENTION: TAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE  
; FILE REFERENCE: 36689.8  
; CURRENT APPLICATION NUMBER: US/11/070.627  
; CURRENT FILING DATE: 2005-03-02  
; PRIOR APPLICATION NUMBER: 60/549,399  
; PRIOR FILING DATE: 2004-03-02  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 4  
; LENGTH: 750  
; TYPE: PRT  
; ORGANISM: Oryctolagus cuniculus  
US-11-070-627-4

Query Match 9.2%; Score 78; DB 7; Length 750;  
Best Local Similarity 23.4%; Pred. No. 3.4;  
Matches 44; Conservative 25; Mismatches 41; Indels 78; Gaps 11;  
QY 1 MPKTIYFGAGWFTDRQNKA YKAEALKEKNTTIDLENSYVPLDNOYKGI RVD E H----- 55  
Db 236 LPSRDYBCTG-----IYKEACTAY-----VDFMLAVAKLIRQEGLPIDENQISVE 272  
QY 56 -----PEYLDKWKATATYNN-----DLNGIK-----TNDIM- 82  
Db 273 MNKVMLEKEIKANATTKSEDRNDPMLLYNKM TLAQIQNFSLEINGKPFWSNFTNEIMS 332  
QY 83 -LGVIYPIEDVDGLGMEIGYALSQGYVLLVLPDSDYGPKI-----NLMWG-VS 130  
Db 333 TWINIFNEEDV-----VYYA---PEYLIKL-----KPILTKYFPRDFQNLFSWRFIM 377  
QY 131 DNVIKMSQ 138  
Db 378 DLVSSLSR 385

## RESULT 2

US-10-793-626-1948  
; Sequence 1948, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN



Matches	32;	Conservative	16;	Mismatches	52;	Indels	29;	Gaps	6;
Qy	22	KEAMEALKENPTID---	LENSVPLD	NQYKGI	RV-----	DEHPYL	HDK	QW	ATATYN 70
Db	157	KGLEETIKENPTIKAVILNPTN	GTGVEY	SEDEIK	AKAKV	IK	NH	LYVIT	DEIYSTLYT- 215
Qy	71	NDLNGIKTNDIM----	LGVI	YP--	DEEDVGL	GMEL	GV	AL	SGQKYVLLVITPDEYDYGKPIN 123
Db	216	---GVKHFS	IASL	IPER	AIYISGL	SKSHAM	TGYRL	GV	VAGPAKIM-----AEIGKVHG 265
Qy	124	LMSWGVSDN	132						
Db	266	LMVTIT	TD	274					

```

RESULT 6
US-11-070-627-2
; Sequence 2, Application US/11070627
; Publication No. US20050271625A1
; GENERAL INFORMATION:
; APPLICANT: Nash , Kevin R.
; APPLICANT: Burger, Corinna
; TITLE OF INVENTION: TAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 36689.8
; CURRENT APPLICATION NUMBER: US/11/070,627
; CURRENT FILING DATE: 2005-03-02
; PRIOR APPLICATION NUMBER: 60/549,399
; PRIOR FILING DATE: 2004-03-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-11-070-627-2

```

Query Match	8.2%;	Score 69.5;	DB 7;	Length 750;
Best Local Similarity	23.5%;	Pred. NO. 23;		
Matches	47;	Conservative 29;	Mismatches 55;	Indels 69; Gaps 12;
Qy	1	MPKKTIVFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNQYKIGRVDEH-----	55	
Db	226	LPSRDYECTG-----IYKEACTAY-----VDPMISVARLFEQRLPIDENQLSLE	272	
Qy	56	-----PYYLHKDWATATYNN-----DLNGIK-----TNDIM-	82	
Db	273	MNKVMELEKEIATATTKPEDRNDPMLLYNKYMTLAKLQNNFSLSEINGKPFWSNFTNIMS	332	
Qy	83	-LGVIYIPDEEDVGLGMELGYA---LSQGYVLLVTPDEDYCKPINLMSWG-VSDNVIVKMS	137	
Db	333	TWNINIQEEEEV---VVYAPYLTUKLPILTKYSPRDLQ---NLMSRRFIMDLVSSLS	384	
Qy	138	QLKDFNFNFKRPDPFEGAVY	157	
Db	185	R---NYKESR-NAPRKALY	399	

```

RESULT 7
US-11-070-627-9
; Sequence 9, Application US/11070627
; Publication No. US20050271625A1
; GENERAL INFORMATION:
; APPLICANT: Nash , Kevin R.
; ATTORNEY: Burger, Corinna
; TITLE OF INVENTION: rAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 36689.8
; CURRENT APPLICATION NUMBER: US/11/070,627
; CURRENT FILING DATE: 2005-03-02
; PRIOR APPLICATION NUMBER: 60/549,399
; PRIOR FILING DATE: 2004-03-02
; NUMBER OF SEQ ID NOS: 11
; SEQUENCE: Patent in version 3.3
; SEQ ID NO 9

```

```

; LENGTH: 786
; TYPE: PRT
; ORGANISM: Canis familiaris
US-11-070-627-9

      8.2%; Score 69.5; DB 7; Length 786;
Query Match Best Local Similarity 24.9%; Pred.No. 24;
Matches 42; Conservative 28; Mismatches 58; Indels 41; Gaps 11.

Qy   9  GAGWFTDRONKAYKEAMEALKENPTIDLENSYV-----PLDNQYKGIRVDEHPHYLVHDKV 63
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db   288 GNGLIRKEKGLLIDENQLSLEMNRVMLEKEIASATTKPEDR-----NDPMLLYNKM 339

Qy   64  WATATYNN-----DLNGIK-----TNDTM--LGUVIPDEEDVGLOMELGYA---LSQGKYV 109
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db   340 TLAIQIQRNFTELDIGDFPSWSNFTEMTSNINIPNEEEV-----VVYAPEYLTUKLI 394

Qy   110 LLVIPDDYGKPINLMSWG-VSDNVIKMSQLKDFNFKNRPDFFEGAVY 157
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db   395 LTKYSSRDLO---NLMSWRITMDLVSLSR-----NYKESR-NAPRKALY 435

```

```

RESULT 8
US-10-793-626-668
; Sequence 668, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 668
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-668

```

Query Match	8.1%; Score 68.5; DB 6; Length 395;
Best Local Similarity	18.1%; Pred. No. 12;
Matches	41; Conservative 33; Mismatches 58; Indels 95; Gaps 10
QY	14 TDRN-KAYKEAMBALK-----ENPTIDLENS---YVPLDNOVKGIRVDEH 55
Db	179 TPRNKEITYENATSETPKDLNDPKQATASVTKPTLEYQSTNEMAYEVNQLKNFRIET 238
QY	56 PEY-----LHDKVWATATYN----- 70
Db	239 QGYNSWCAGYTMSALFNATYNTNRYNABSMRYLHPNLRGHDFQFTGLTSMELRFGRSQ 298
QY	71 -----NDLNGIKTND---IMLGVIYPDEEDVGLGMEIGYA---LSOGKYVL 110
Db	299 GRNTQYLNRMSTSYNEVDQTLTNNOGIAVLGRKRVESDGIHACHAMAVAGNKNVNGQKVI 358
QY	111 LVIPDEEDYGGKPINLMSVGSDNVIKMSQLDKDFNPNKPRDFYEGAVY 157
Db	359 LWNPDWNG----LMTQDAHNIIPVS-----NGDHYEWY-ASIIY 393

RESULT 9  
US-10-485-517-145  
; Sequence 145, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated

```

; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 145
; LENGTH: 1107
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-145

Query Match      8.1%; Score 68.5; DB 6; Length 1107;
Best Local Similarity 20.4%; Pred. No. 48;
Matches 37; Conservative 26; Mismatches 65; Indels 53; Gaps 6;

QY 17 QNKAYKEAMEALKEN-----PTIDLENSYVPLDNOYKGI----- 50
DB 618 QAQALNEAMKALKESIKDQFQTEASSKFINEQQAQDAYTQAVQHAKDLINKTTDPTLAK 677
QY 51 -----RVDEHPEYLH-----DKWATATYNNDLNGIKTNDIMLGVVYIPDEEDVGL 95
DB 678 SIIDQATQAVTDKNNLHGQKLAQDKQRATETLNN-LSNLT-----PQRQ--AL 725
QY 96 GMELGYALSQGYKVVLLVIPPEDYKPINLMSWGVSDNVKMSQKIDPFNFKPRDFYEGA 155
DB 726 ENQINNAATRGVAQKLTEAQAALNQAMEALRNSIQDQOQTEAGSKFINEDKPKQDAYQAA 785
QY 156 V 156
DB 786 V 786

RESULT 10
US-11-094-586-4
; Sequence 4, Application US/11094586
; Publication No. US20050273886A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Butler, Karlene H.
; APPLICANT: Carlson, Thomas J.
; APPLICANT: Hitz, William D.
; APPLICANT: Scoop, Johan M.
; TITLE OF INVENTION: Plastidic Phosphoglucosyltransferase Genes
; FILE REFERENCE: BB1451 USCIP
; CURRENT APPLICATION NUMBER: US/11/094,586
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: US 09/906,209
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: US 60/218,712
; PRIOR FILING DATE: 2000-07-17
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97 & PatentIn Version 3.3
; SEQ ID NO 4
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Zea mays
US-11-094-586-4

Query Match      8.0%; Score 68; DB 7; Length 605;
Best Local Similarity 29.5%; Pred. No. 24;
Matches 31; Conservative 17; Mismatches 29; Indels 28; Gaps 8;

QY 56 PEYLHDKWATATYNNDL--NGIKTNDIMLGVIIPDEEDVGLGMELGYALSQGYKVVLLVI 113
DB 186 PETITDQI-----YGNLTLSISEIKTAD-----IPDIDLSSVG-----VSYGDFAIEVI 229

; APPLICANT: 114 -PDYDYKPINLMSWGVSDNVKMSQKIDPFNFKPRDFYEGAVY 157
; APPLICANT: 230 DPVSDY---LELM-----ENVDFQLIKDL-LSRPDRFEITFDAMH 265
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 145
; LENGTH: 1107
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-145

Query Match      8.1%; Score 68.5; DB 6; Length 1107;
Best Local Similarity 20.4%; Pred. No. 48;
Matches 37; Conservative 26; Mismatches 65; Indels 53; Gaps 6;

QY 17 QNKAYKEAMEALKEN-----PTIDLENSYVPLDNOYKGI----- 50
DB 618 QAQALNEAMKALKESIKDQFQTEASSKFINEQQAQDAYTQAVQHAKDLINKTTDPTLAK 677
QY 51 -----RVDEHPEYLH-----DKWATATYNNDLNGIKTNDIMLGVVYIPDEEDVGL 95
DB 678 SIIDQATQAVTDKNNLHGQKLAQDKQRATETLNN-LSNLT-----PQRQ--AL 725
QY 96 GMELGYALSQGYKVVLLVIPPEDYKPINLMSWGVSDNVKMSQKIDPFNFKPRDFYEGA 155
DB 726 ENQINNAATRGVAQKLTEAQAALNQAMEALRNSIQDQOQTEAGSKFINEDKPKQDAYQAA 785
QY 156 V 156
DB 786 V 786

RESULT 11
US-10-821-234-1662
; Sequence 1662, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1662
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1662

Query Match      8.0%; Score 68; DB 6; Length 747;
Best Local Similarity 38.2%; Pred. No. 32;
Matches 13; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

QY 114 PDYDYKPINLMSWGVSDNVKMSQKIDPFNFKP 147
DB 612 PDEIQRPVVRVPSWGLEDNVVCSPAR--NFSRP 643

RESULT 12
US-11-070-627-1
; Sequence 1, Application US/11070627
; Publication No. US20050271625A1
; GENERAL INFORMATION:
; APPLICANT: Nash, Kevin R.
; APPLICANT: Burger, Corinna
; TITLE OF INVENTION: rAAV-NEPRILYSIN COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 36689.8
; CURRENT APPLICATION NUMBER: US/11/070,627
; CURRENT FILING DATE: 2005-03-02
; PRIOR APPLICATION NUMBER: 60/549,399
; PRIOR FILING DATE: 2004-03-02
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-070-627-1

Query Match      8.0%; Score 68; DB 7; Length 750;
Best Local Similarity 24.3%; Pred. No. 32;
Matches 44; Conservative 23; Mismatches 50; Indels 64; Gaps 10;

QY 1 MPKKTIVFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGVIRVDEH----- 55
DB 226 LPSRDYECTG-----IYKACTAY-----VDFMISVARLIRQERLPIDENQALALE 272
QY 56 -----PEYLHDKWATATYNN---DLNGIK-----TNDIM- 82
DB 273 MNKVMELEKEIANATAKPEDRNDPMLLYNKMRLAQIQQNNFSLSEINGKPFSLNFTNEIMS 332
QY 83 -LGVYIPDEEDVGLGMELGYA---LSQGYKVVLLVIPPEDYKGINLMSWG-VSDNVKMS 137
DB 333 TVNISITNEEDV-----VYAPYELTKLPILTYSARDLQ---NLSMWRFMIDVSSLS 384
```

Qy 138 Q 138  
Db 385 R 385

## RESULT 13

US-10-821-234-1007  
; Sequence 1007, Application US/10821234  
; Publication No. US20050255114A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; CURRENT FILING DATE: 2004-04-07  
; PRIOR APPLICATION NUMBER: US 60/462,047  
; PRIOR FILING DATE: 2003-04-07  
; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: pt SEQ\_genes Version 1.0  
; SEQ ID NO 1007  
; LENGTH: 751  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-821-234-1007

Query Match 8.0%; Score 68; DB 6; Length 751;  
Best Local Similarity 24.3%; Pred. No. 32;  
Matches 44; Conservative 23; Mismatches 50; Indels 64; Gaps 10;  
Qy 1 MPKTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKQIRVDEH-----55  
Db 227 LPSRDYECTG-----IYKEACTAY-----VDFMISVARLIRQBERLPIDENQALE 273  
Qy 56 -----PEYLDKQWATATYNN-----DLNGIK-----TNDIM- 82  
Db 274 MNKVMLEKEIANATAKPEDRNDPMLYNNKRLAQIQQNFSLEINGKPFSLNFTNEIMS 333  
Qy 83 -LGVIYPERDVGLGMEIGYA---LSQGYVLLVLPDEDYDKPINLMSWG-VSDNVVIMS 137  
Db 334 TVNISITNEEDV-----VVIYAPEYLTUKLPILTKYSARDLQ---NLMSWRFIMDLVSSLS 385  
Qy 138 Q 138  
Db 386 R 386

## RESULT 14

US-11-004-057-6  
; Sequence 6, Application US/11004057  
; Publication No. US20050244846A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Gary L.  
; TITLE OF INVENTION: MEK1 PROTEINS AND FRAGMENTS THEREOF FOR USE IN REGULATING  
; FILE REFERENCE: APOPTOSIS  
; CURRENT APPLICATION NUMBER: US/11/004,057  
; CURRENT FILING DATE: 2004-12-02  
; PRIOR APPLICATION NUMBER: US/09/403,075  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 1302  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-004-057-6

Query Match 8.0%; Score 67.5; DB 7; Length 1302;  
Best Local Similarity 23.9%; Pred. No. 74;

Matches 37; Conservative 21; Mismatches 62; Indels 35; Gaps 6;  
Qy 18 NKAYKEAMEALKEN-----PTIDLENSYVPLDNOYKQIRVDEHPEYL 59  
Db 964 NQCKEKMEAEAEALAIAMAMSASQDALPIVPOLOQVENGEDII-----IIQDITPETL 1017  
Qy 60 HDKVMATATYNNDLNGIKTNDIMLGVIYIP--DEEDVGLGMEIGYALSQGYVLLVLPDED 117  
Db 1018 PGHTKAKQPYREDAEWLKGQOIGLGAFFSSCYQAQDVGTGLM--AVKQVTVRNTSSQEQ 1075  
Qy 118 -----YGKPINLMSWGVSDNVIKM--SOLKDFNFN 145  
Db 1076 EVVEALREERIMMSHLNPNIIRMLGATCEKSNYN 1110

## RESULT 15

US-11-004-057-21  
; Sequence 21, Application US/11004057  
; Publication No. US20050244846A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Gary L.  
; TITLE OF INVENTION: MEK1 PROTEINS AND FRAGMENTS THEREOF FOR USE IN REGULATING  
; FILE REFERENCE: APOPTOSIS  
; CURRENT APPLICATION NUMBER: US/11/004,057  
; CURRENT FILING DATE: 2004-12-02  
; PRIOR APPLICATION NUMBER: US/09/403,075  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 1493  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-11-004-057-21

Query Match 8.0%; Score 67.5; DB 7; Length 1493;  
Best Local Similarity 24.5%; Pred. No. 89;  
Matches 38; Conservative 20; Mismatches 62; Indels 35; Gaps 7;  
Qy 18 NKAYKEAMEALKEN-----PTIDLENSYVPLDNOYKQIRVDEHPEYL 59  
Db 1155 NQCKEKMEAEAEALAIAMAMSASQDALPIVPOLOQVENG-----EDIIIIQADT-PETL 1208  
Qy 60 HDKVMATATYNNDLNGIKTNDIMLGVIYIP--DEEDVGLGMEIGYALSQGYVLLVLPDED 117  
Db 1209 PGHTKANEPYREDTEWLKQOIGLGAFFSSCYQAQDVGTGLM--AVKQVTVRNTSSQEQ 1266  
Qy 118 -----YGKPINLMSWGVSDNVIKM--SOLKDFNFN 145  
Db 1267 EVVEALREERIMMSHLNPNIIRMLGATCEKSNYN 1301

Search completed: December 23, 2005, 22:20:50  
Job time : 19 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:57:09 ; Search time 569 Seconds  
(without alignments)  
381.313 Million cell updates/sec

Title: US-10-049-750-14  
Perfect score: 848  
Sequence: 1 MPKTIYFGAGWFTDRQNK.....OLKDFNFKPRDFYEGAVY 157

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 7861189 seqs, 1381955077 residues

Total number of hits satisfying chosen parameters: 7861189

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA Main:\*

- 1: /cgn2\_6/ptodata/1/paa/PCTUS COMB.pep:\*\*
- 2: /cgn2\_6/ptodata/1/paa/US066 COMB.pep:\*\*
- 3: /cgn2\_6/ptodata/1/paa/US073 COMB.pep:\*\*
- 4: /cgn2\_6/ptodata/1/paa/US074 COMB.pep:\*\*
- 5: /cgn2\_6/ptodata/1/paa/US075 COMB.pep:\*\*
- 6: /cgn2\_6/ptodata/1/paa/US076 COMB.pep:\*\*
- 7: /cgn2\_6/ptodata/1/paa/US077 COMB.pep:\*\*
- 8: /cgn2\_6/ptodata/1/paa/US078 COMB.pep:\*\*
- 9: /cgn2\_6/ptodata/1/paa/US079 COMB.pep:\*\*
- 10: /cgn2\_6/ptodata/1/paa/US080 COMB.pep:\*\*
- 11: /cgn2\_6/ptodata/1/paa/US081 COMB.pep:\*\*
- 12: /cgn2\_6/ptodata/1/paa/US082 COMB.pep:\*\*
- 13: /cgn2\_6/ptodata/1/paa/US083 COMB.pep:\*\*
- 14: /cgn2\_6/ptodata/1/paa/US084 COMB.pep:\*\*
- 15: /cgn2\_6/ptodata/1/paa/US085 COMB.pep:\*\*
- 16: /cgn2\_6/ptodata/1/paa/US086 COMB.pep:\*\*
- 17: /cgn2\_6/ptodata/1/paa/US087 COMB.pep:\*\*
- 18: /cgn2\_6/ptodata/1/paa/US088 COMB.pep:\*\*
- 19: /cgn2\_6/ptodata/1/paa/US089 COMB.pep:\*\*
- 20: /cgn2\_6/ptodata/1/paa/US090 COMB.pep:\*\*
- 21: /cgn2\_6/ptodata/1/paa/US091 COMB.pep:\*\*
- 22: /cgn2\_6/ptodata/1/paa/US092 COMB.pep:\*\*
- 23: /cgn2\_6/ptodata/1/paa/US093 COMB.pep:\*\*
- 24: /cgn2\_6/ptodata/1/paa/US094 COMB.pep:\*\*
- 25: /cgn2\_6/ptodata/1/paa/US095 COMB.pep:\*\*
- 26: /cgn2\_6/ptodata/1/paa/US096 COMB.pep:\*\*
- 27: /cgn2\_6/ptodata/1/paa/US097 COMB.pep:\*\*
- 28: /cgn2\_6/ptodata/1/paa/US098 COMB.pep:\*\*
- 29: /cgn2\_6/ptodata/1/paa/US099 COMB.pep:\*\*
- 30: /cgn2\_6/ptodata/1/paa/US100 COMB.pep:\*\*
- 31: /cgn2\_6/ptodata/1/paa/US101 COMB.pep:\*\*
- 32: /cgn2\_6/ptodata/1/paa/US102 COMB.pep:\*\*
- 33: /cgn2\_6/ptodata/1/paa/US103 COMB.pep:\*\*
- 34: /cgn2\_6/ptodata/1/paa/US104 COMB.pep:\*\*
- 35: /cgn2\_6/ptodata/1/paa/US105 COMB.pep:\*\*
- 36: /cgn2\_6/ptodata/1/paa/US106 COMB.pep:\*\*
- 37: /cgn2\_6/ptodata/1/paa/US107 COMB.pep:\*\*
- 38: /cgn2\_6/ptodata/1/paa/US108 COMB.pep:\*\*
- 39: /cgn2\_6/ptodata/1/paa/US109 COMB.pep:\*\*
- 40: /cgn2\_6/ptodata/1/paa/US110 COMB.pep:\*\*
- 41: /cgn2\_6/ptodata/1/paa/US111 COMB.pep:\*\*
- 42: /cgn2\_6/ptodata/1/paa/US112 COMB.pep:\*\*
- 43: /cgn2\_6/ptodata/1/paa/US114 COMB.pep:\*\*

- 44: /cgn2\_6/ptodata/1/paa/US600 COMB.pep:\*\*
- 45: /cgn2\_6/ptodata/1/paa/US601 COMB.pep:\*\*
- 46: /cgn2\_6/ptodata/1/paa/US602 COMB.pep:\*\*
- 47: /cgn2\_6/ptodata/1/paa/US603 COMB.pep:\*\*
- 48: /cgn2\_6/ptodata/1/paa/US604 COMB.pep:\*\*
- 49: /cgn2\_6/ptodata/1/paa/US605 COMB.pep:\*\*
- 50: /cgn2\_6/ptodata/1/paa/US606 COMB.pep:\*\*
- 51: /cgn2\_6/ptodata/1/paa/US607 COMB.pep:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	848	100.0	157	30	US-10-049-750-14	Sequence 14, Appl
2	728	85.8	158	40	US-11-097-292-2	Sequence 2, Appl
3	728	85.8	158	40	US-11-097-292-14	Sequence 14, Appl
4	719	84.8	159	40	US-11-097-292-12	Sequence 12, Appl
5	719	84.8	159	50	US-60-622-712-264	Sequence 264, App
6	719	84.8	159	51	US-60-711-491-264	Sequence 264, App
7	629	74.2	133	40	US-11-097-292-10	Sequence 10, Appl
8	620	73.1	165	27	US-09-791-537-8009	Sequence 8009, Ap
9	405	47.8	84	40	US-11-097-292-8	Sequence 8, Appl
10	209	24.6	168	40	US-11-097-292-6	Sequence 6, Appl
11	166.5	19.6	167	40	US-11-097-292-4	Sequence 4, Appl
12	154.5	18.2	167	50	US-60-622-712-2174	Sequence 2174, Ap
13	154.5	18.2	167	51	US-60-711-491-2174	Sequence 2174, Ap
14	91.5	10.8	156	34	US-10-417-884A-6487	Sequence 6487, Ap
15	91.5	10.8	156	34	US-10-417-884A-6487	Sequence 6487, Ap
16	91	10.7	719	27	US-09-791-537-10210	Sequence 10210, A
17	91	10.7	719	27	US-09-791-537-96101	Sequence 96101, A
18	85.5	10.1	438	26	US-09-629-469A-18384	Sequence 18384, A
19	85.5	10.1	438	39	US-10-917-503-18384	Sequence 18384, A
20	85.5	10.1	456	34	US-10-455-772-22	Sequence 22, Appl
21	85.5	10.1	464	1	PCT-US00-04340-48	Sequence 48, Appl
22	85.5	10.1	464	25	US-09-507-209-48	Sequence 48, Appl
23	85.5	10.1	464	28	US-09-808-602-95	Sequence 95, Appl
24	85.5	10.1	487	28	US-09-800-198-80	Sequence 80, Appl
25	85.5	10.1	487	28	US-09-808-602-93	Sequence 93, Appl
26	85.5	10.1	487	34	US-10-455-772-16	Sequence 16, Appl
27	85.5	10.1	487	34	US-10-455-772-46	Sequence 46, Appl
28	85.5	10.1	487	34	US-10-455-772-48	Sequence 48, Appl
29	85.5	10.1	487	34	US-10-455-772-50	Sequence 50, Appl
30	85.5	10.1	487	34	US-10-455-772-52	Sequence 52, Appl
31	85.5	10.1	487	38	US-10-844-295-80	Sequence 80, Appl
32	84.5	10.0	198	45	US-60-173-469-1175	Sequence 1175, Ap
33	84.5	10.0	198	45	US-60-198-955-1449	Sequence 1449, Ap
34	84.5	10.0	243	34	US-10-455-772-38	Sequence 38, Appl
35	84.5	10.0	247	34	US-10-455-772-20	Sequence 20, Appl
36	84.5	10.0	396	1	PCT-US01-11988-2176	Sequence 2176, Ap
37	84.5	10.0	396	28	US-09-833-245-2176	Sequence 2176, Ap
38	84.5	10.0	396	28	US-09-833-245A-2176	Sequence 2176, Ap
39	84.5	10.0	396	28	US-09-833-245B-2176	Sequence 86, Appl
40	84.5	10.0	396	34	US-10-458-714-86	Sequence 86, Appl
41	84.5	10.0	452	28	US-09-808-602-21	Sequence 21, Appl
42	84.5	10.0	452	34	US-10-455-772-42	Sequence 42, Appl
43	84.5	10.0	456	34	US-10-455-772-18	Sequence 18, Appl
44	84.5	10.0	456	34	US-10-455-772-24	Sequence 24, Appl
45	84.5	10.0	456	34	US-10-455-772-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1  
US-10-049-750-14  
; Sequence 14, Application US/10049750  
; GENERAL INFORMATION:  
; APPLICANT: Tischer, Wilhelm

```
; APPLICANT: Ihlenfeldt, Hans-Georg
; APPLICANT: Barzu, Octavian
; APPLICANT: Sakamoto, Hiroshi
; APPLICANT: Pistotnik, Blisabeth
; APPLICANT: Marliere, Philippe
; APPLICANT: Pochet, Sylvie
; TITLE OF INVENTION: Enzymatic synthesis of deoxyribonucleosides
; FILE REFERENCE: 20373PWO Deoxyribonucleosides
; CURRENT APPLICATION NUMBER: US/10/049,750
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: EP99116425.2
; PRIOR FILING DATE: 1999-08-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 14
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Lactobacillus leichmannii
US-10-049-750-14

Query Match      100.0%; Score 848; DB 30; Length 157;
Best Local Similarity 100.0%; Pred. No. 7.3e-87;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPKTTYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLH 60
DB 1 MPKTTYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLH 60

QY 61 DKWATATYNDLNGIKTNDIMLGVIYIPDEBDVGLGMEYALSGQKYLIVIPDEYDK 120
DB 61 DKWATATYNDLNGIKTNDIMLGVIYIPDEBDVGLGMEYALSGQKYLIVIPDEYDK 120

QY 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 121 PINLMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157

RESULT 2
US-11-097-292-2
; Sequence 2, Application US/11097292
; GENERAL INFORMATION:
; APPLICANT: KAMINSKI, PIERRE-ALEXANDRE
; APPLICANT: TAILLIEZ, PATRICK
; APPLICANT: MARLIERE, PHILIPPE
; APPLICANT: QUENEZ, PASCAL
; APPLICANT: COTAYA, RACHEL
; TITLE OF INVENTION: LACTOBACILLUS N-DEOXYRIBOSYL TRANSFERASES, CORRESPONDING
; FILE REFERENCE: 250246USOXPT
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: US/11/097,292
; PRIOR FILING DATE: 2001-09-14
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Lactobacillus helveticus NTD
US-11-097-292-2

Query Match      85.8%; Score 728; DB 40; Length 158;
Best Local Similarity 82.6%; Pred. No. 2.8e-73;
Matches 128; Conservative 20; Mismatches 7; Indels 0; Gaps 0;

QY 3 KKTIFYGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLHDK 62
DB 4 KKTLYFGAGWNEKQNKAYKEAMEALKENPTVDLENSYVPLENQYKGRIDEHPEYLHNI 63

QY 63 WATATYNDLNGIKTNDIMLGVIYIPDEBDVGLGMEYALSGQKYLIVIPDEYDKPI 122
DB 64 EWASATYHNDLVGIKTSDVNLGVYLPEEDVGLGMEYALSGQKYLIVIPDEYDKPI 123

QY 123 NLMMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 124 NLMMSGVCDNAIKISELKDFDNKPRYFNFDGAVY 158

RESULT 3
US-11-097-292-14
; Sequence 14, Application US/11097292
; GENERAL INFORMATION:
; APPLICANT: KAMINSKI, PIERRE-ALEXANDRE
; APPLICANT: TAILLIEZ, PATRICK
; APPLICANT: MARLIERE, PHILIPPE
; APPLICANT: QUENEZ, PASCAL
; APPLICANT: COTAYA, RACHEL
; TITLE OF INVENTION: LACTOBACILLUS N-DEOXYRIBOSYL TRANSFERASES, CORRESPONDING
; FILE REFERENCE: 250246USOXPT
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: US/11/097,292
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/FR02/03120
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 14
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Lactobacillus helveticus NTD
US-11-097-292-14

Query Match      85.8%; Score 728; DB 40; Length 158;
Best Local Similarity 82.6%; Pred. No. 2.8e-73;
Matches 128; Conservative 20; Mismatches 7; Indels 0; Gaps 0;

QY 3 KKTIFYGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLHDK 62
DB 4 KKTLYFGAGWNEKQNKAYKEAMEALKENPTVDLENSYVPLENQYKGRIDEHPEYLHNI 63

QY 63 WATATYNDLNGIKTNDIMLGVIYIPDEBDVGLGMEYALSGQKYLIVIPDEYDKPI 122
DB 64 EWASATYHNDLVGIKTSDVNLGVYLPEEDVGLGMEYALSGQKYLIVIPDEYDKPI 123

QY 123 NLMMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 124 NLMMSGVCDNAIKISELKDFDNKPRYFNFDGAVY 158

RESULT 4
US-11-097-292-12
; Sequence 12, Application US/11097292
; GENERAL INFORMATION:
; APPLICANT: KAMINSKI, PIERRE-ALEXANDRE
; APPLICANT: TAILLIEZ, PATRICK
; APPLICANT: MARLIERE, PHILIPPE
; APPLICANT: QUENEZ, PASCAL
; APPLICANT: COTAYA, RACHEL
; TITLE OF INVENTION: LACTOBACILLUS N-DEOXYRIBOSYL TRANSFERASES, CORRESPONDING
; FILE REFERENCE: 250246USOXPT
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: US/11/097,292
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/FR02/03120
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 12
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Lactobacillus helveticus NTD
US-11-097-292-12

Query Match      85.8%; Score 728; DB 40; Length 158;
Best Local Similarity 82.6%; Pred. No. 2.8e-73;
Matches 128; Conservative 20; Mismatches 7; Indels 0; Gaps 0;

QY 3 KKTIFYGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDEHPEYLHDK 62
DB 4 KKTLYFGAGWNEKQNKAYKEAMEALKENPTVDLENSYVPLENQYKGRIDEHPEYLHNI 63

QY 63 WATATYNDLNGIKTNDIMLGVIYIPDEBDVGLGMEYALSGQKYLIVIPDEYDKPI 122
DB 64 EWASATYHNDLVGIKTSDVNLGVYLPEEDVGLGMEYALSGQKYLIVIPDEYDKPI 123

QY 123 NLMMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 124 NLMMSGVCDNAIKISELKDFDNKPRYFNFDGAVY 158
```

```
; ORGANISM: Lactobacillus acidophilus NTD
US-11-097-292-12

Query Match      84.8%; Score 719; DB 40; Length 159;
Best Local Similarity 83.1%; Pred. No. 2.9e-72;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 4 KTIYFGAGWFTDRQNKAYKAEALKENPTTIDLENSYVPLDNOYKGRVDEHPEYLDHVKV 63
DB 6 KTLIFGAGWGFNEKQNKAYKAEALKQNPVDLENSYVPLENQYKDIRVDEHPEYLDHIE 65
QY 64 WATATYNNDLNGIKTNDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 123
DB 66 WASATYHNDLIGIKSSDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 125
QY 124 LMSGVSDNVIRKMSQLKDFNFKPRDFYEGAVY 157
DB 126 LMSGVCDNAIKISELKDFDNFKPRFNFDGAVY 159

RESULT 5
US-60-622-712-264
; Sequence 264, Application US/60622712
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd
; APPLICANT: Russell, William
; APPLICANT: Alterman, Eric
; TITLE OF INVENTION: LACTOBACILLUS ACIDOPHILUS NUCLEIC ACIDS AND USES THEREOF
; FILE REFERENCE: 5051-604PR10
; CURRENT APPLICATION NUMBER: US/60/622,712
; CURRENT FILING DATE: 2004-10-27
; NUMBER OF SEQ ID NOS: 2559
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 264
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-60-622-712-264

Query Match      84.8%; Score 719; DB 50; Length 159;
Best Local Similarity 83.1%; Pred. No. 2.9e-72;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 4 KTIYFGAGWFTDRQNKAYKAEALKENPTTIDLENSYVPLDNOYKGRVDEHPEYLDHVKV 63
DB 6 KTLIFGAGWGFNEKQNKAYKAEALKQNPVDLENSYVPLENQYKDIRVDEHPEYLDHIE 65
QY 64 WATATYNNDLNGIKTNDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 123
DB 66 WASATYHNDLIGIKSSDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 125
QY 124 LMSGVSDNVIRKMSQLKDFNFKPRDFYEGAVY 157
DB 126 LMSGVCDNAIKISELKDFDNFKPRFNFDGAVY 159

RESULT 6
US-60-711-491-264
; Sequence 264, Application US/60711491
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd
; APPLICANT: Russell, William
; APPLICANT: Altermann, Eric
; TITLE OF INVENTION: LACTOBACILLUS ACIDOPHILUS NUCLEIC ACIDS AND USES THEREOF
; FILE REFERENCE: 035051/296336
; CURRENT APPLICATION NUMBER: US/60/711,491
; CURRENT FILING DATE: 2005-08-26
; NUMBER OF SEQ ID NOS: 2561
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 264
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
```

```
US-60-711-491-264

Query Match      84.8%; Score 719; DB 51; Length 159;
Best Local Similarity 83.1%; Pred. No. 2.9e-72;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 4 KTIYFGAGWFTDRQNKAYKAEALKENPTTIDLENSYVPLDNOYKGRVDEHPEYLDHVKV 63
DB 6 KTLIFGAGWGFNEKQNKAYKAEALKQNPVDLENSYVPLENQYKDIRVDEHPEYLDHIE 65
QY 64 WATATYNNDLNGIKTNDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 123
DB 66 WASATYHNDLIGIKSSDIMLGVIIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPIN 125
QY 124 LMSGVSDNVIRKMSQLKDFNFKPRDFYEGAVY 157
DB 126 LMSGVCDNAIKISELKDFDNFKPRFNFDGAVY 159

RESULT 7
US-11-097-292-10
; Sequence 10, Application US/11097292
; GENERAL INFORMATION:
; APPLICANT: KAMINSKI, PIERRE-ALEXANDRE
; APPLICANT: TAILLIEZ, PATRICK
; APPLICANT: MARLIERE, PHILIPPE
; APPLICANT: QUENEZ, PASCAL
; APPLICANT: COTAYA, RACHEL
; TITLE OF INVENTION: LACTOBACILLUS N-DEOXYRIBOSYL TRANSFERASES, CORRESPONDING
; FILE REFERENCE: 250246US0XPC7
; CURRENT APPLICATION NUMBER: US/11/097,292
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: FR 0 111 911
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/FR02/03120
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 10
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Lactobacillus amylovorus NTD
US-11-097-292-10

Query Match      74.2%; Score 629; DB 40; Length 133;
Best Local Similarity 84.2%; Pred. No. 3.5e-62;
Matches 112; Conservative 17; Mismatches 4; Indels 0; Gaps 0;

QY 25 MEALKENPTIDLENSYVPLDNOYKGRVDEHPEYLDHVKWATATYNNDLNGIKTNDIMLG 84
DB 1 MEALKKNPTVDLENSYVPLDNOYKGRVDEHPEYLDHIEWASSTYHNDLVGKSSDVMUG 60
QY 85 VYIPDEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPINLMSGVSDNVIRKMSQLKDFNF 144
DB 61 VYLPEEDVGLGMELGVALSGQKYVLLVIPDEDYGKPINLMSGVCDNVIKISELKDFDF 120
QY 145 NKPRFDFYEGAVY 157
DB 121 NRPRFNFDGAVY 133

RESULT 8
US-09-791-537-8009
; Sequence 8009, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
```





```
RESULT 15
US-10-417-884A-6487
; Sequence 6487, Application US/10417884A
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/417,884A
; FILING DATE: 17-Apr-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6487:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...156
; SEQUENCE DESCRIPTION: SEQ ID NO: 6487:
US-10-417-884A-6487

Query Match      10.8%; Score 91.5; DB 34; Length 156;
Best Local Similarity 23.0%; Pred. No. 0.28;
Matches 35; Conservative 28; Mismatches 74; Indels 15; Gaps 5;

Qy      6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNQYKGIRVDHEHPEYLHDKVWA 65
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      14 IYLAGPFFSEQIDRVSRIEKALEENKTIV--TSFYSPRHQ-----ESNYELFSAGWA 64
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Qy      66 TATYNDNLNGIKTNDIMLGVIYIDEE--DVGLGMELGYALSQGYVLLVIPDEDYGGPIN 123
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      65 QEYVEKOMEELTNAEYVAILDFEHQIDPGTAYELGVATMLKPKMIIV---QEETVPTN 121
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Qy      124 LMSGVSDNVIKMSQ-LKDFNFKPRDFYEG 154
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      122 LMITQSLHTYLYKSDQAVREYDFETLPVETVYG 153
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Search completed: December 23, 2005, 22:15:22  
Job time : 572 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 22:00:24 ; Search time 14 Seconds  
(without alignments)  
118.825 Million cell updates/sec

Title: US-10-049-750-14

Perfect score: 848

Sequence: 1 MPKXITFGAGWFTDRQNK.....QLKDFNFKRFDYEGAVY 157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 62187 seqs, 10595856 residues

Total number of hits satisfying chosen parameters: 62187

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New.\*

- 1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US11\_NEW\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84.5	10.0	396	7	US-11-264-096-2176
2	83	9.8	3354	8	US-60-732-162-1678
3	74	8.7	357	7	US-11-045-004-1742
4	73	8.6	1310	6	US-10-556-060-326
5	72.5	8.5	578	8	US-60-732-162-404
6	72.5	8.5	578	8	US-60-742-219-464
7	72.5	8.5	710	7	US-11-288-493-44
8	72	8.5	893	6	US-10-244-081A-28
9	70.5	8.3	844	7	US-11-127-877A-40
10	70.5	8.3	844	8	US-60-732-162-402
11	70.5	8.3	844	8	US-60-742-219-462
12	70.5	8.3	899	8	US-60-732-162-408
13	70.5	8.3	899	8	US-60-742-219-466
14	70.5	8.3	961	8	US-60-732-162-406
15	70.5	8.3	961	8	US-60-742-219-460
16	70.5	8.3	1576	7	US-11-238-031-5
17	69.5	8.2	192	7	US-11-045-004-607
18	69	8.1	138	7	US-11-045-004-1152
19	69	8.1	743	1	PCT-US05-20516-2
20	69	8.1	749	8	US-60-742-871-285
21	69	8.1	750	8	US-60-742-871-287
22	69	8.1	750	8	US-60-742-871-288
23	69	8.1	750	8	US-60-742-871-289
24	69	8.1	750	8	US-60-742-871-290
25	69	8.1	750	8	US-60-742-871-291

26	69	8.1	750	8	US-60-742-871-292
27	69	8.1	782	1	PCT-US05-20516-4
28	69	8.1	831	8	US-60-742-219-884
29	69	8.1	831	8	US-60-742-871-509
30	69	8.1	831	8	US-60-742-871-511
31	69	8.1	838	8	US-60-742-871-510
32	68	8.0	750	8	US-60-742-871-286
33	67	7.9	721	7	US-11-170-482-12
34	67	7.9	738	7	US-11-045-004-1782
35	66.5	7.8	283	7	US-11-045-004-2575
36	66.5	7.8	580	7	US-11-045-004-938
37	66.5	7.8	779	8	US-60-733-588-11
38	66.5	7.8	779	8	US-60-733-434-11
39	66.5	7.8	902	8	US-60-733-588-9
40	66.5	7.8	902	8	US-60-733-434-9
41	66.5	7.8	1220	1	PCT-US05-11978-6
42	66	7.8	1196	8	US-60-732-162-784
43	65.5	7.7	248	8	US-60-741-048-42
44	65.5	7.7	466	8	US-60-742-872-59
45	65.5	7.7	466	8	US-60-742-872-60

ALIGNMENTS

RESULT 1

US-11-264-096-2176  
; Sequence 2176, Application US/11264096  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: P546D1  
; CURRENT APPLICATION NUMBER: US/11/264,096  
; CURRENT FILING DATE: 2005-11-02  
; PRIOR APPLICATION NUMBER: 09/833,245  
; PRIOR FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/229, 358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 2176  
; LENGTH: 396  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-264-096-2176

Query Match 10.0%; Score 84.5; DB 7; Length 396;  
Best Local Similarity 23.1%; Pred. NO. 0.072;  
Matches 30; Conservative 26; Mismatches 49; Indels 25; Gaps 5;

QY	9	GAGWTFDQNKAYKEAMEALKENPTIDLENSVPLDNOVK-----GIRVDEHPYLDHKV	63
DB	251	GDWLTAAHTTYPKDSVSLRKNQSVNVLGHTADEMLKLGNHVHVVVHPDTRQNS-	309
QY	64	WATATYNNDL-----NGIKTNDIMLGVIYIPDEEDV-----GLGMELGYALSQOK	107
DB	310	--SHNFSGDIALLELOHSIPLGPNVLPVCLPDNETLYRSGLLGYVSGFGMENGWLTTELK	367
QY	108	VYLL-VIPDE 116	
DB	368	YSRLPVAPRE 377	

RESULT 2

US-60-732-162-1678  
; Sequence 1678, Application US/60732162  
; GENERAL INFORMATION:  
; APPLICANT: Belouchi, Abdelmajid  
; APPLICANT: Raelson, John V

```

; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Debrus, Sophie
; APPLICANT: Berdewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated With Asthma Disease
; FILE REFERENCE: 059908-5010-PR
; CURRENT APPLICATION NUMBER: US/60/732,162
; CURRENT FILING DATE: 2005-11-02
; NUMBER OF SEQ ID NOS: 4417
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1678
; LENGTH: 3354
; TYPE: PRT
; ORGANISM: Homosapiens
US-60-732-162-1678

Query Match          9.8%; Score 83; DB 8; Length 3354;
Best Local Similarity 22.2%; Pred. No. 1.9;
Matches 36; Conservative 32; Mismatches 68; Indels 26; Gaps 8;

Qy 15 DRQKVKKEAMEALKE--NPTIDL-----ENSYPVLDNQ-YKGIRVDEHPEYVLDH 61
Db 1593 DRKQSFHLVAIVEDGTPFLSATVHYTVIVDENDNAPMFQPPHYEVLDDEGDFTLNT 1652

Qy 62 KVMATATYNNDD--LNGIKTNDIMLGVIYIPD---BEDVGL---GMELGYALSQGGKYVLLVI 113
Db 1653 SLITIQALDLDEGPGVTVAIVAGNVNTFRIDRHMGVITAAKELDYEISHGRYTLIVT 1712

Qy 114 PDEDYKGPINLMGVSNDVNIKMSQLKDFNFKPRFD-FYEG 154
Db 1713 ATDQ----CPILSHRLTSTTTVLNVNVDINDNVTPFRDYEG 1750

RESULT 3
US-11-045-004-1742
; Sequence 1742, Application US/11045004
; GENERAL INFORMATION:
; APPLICANT: BUCHRIESER, CARMEN
; APPLICANT: FRANGUL, LIONEL
; APPLICANT: COUVE, ELISABETH
; APPLICANT: RUSNIOK, CHRISTOPHE
; APPLICANT: FSIHI, HAFIDA
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DUSSURGET, OLIVIER
; APPLICANT: CHETOJANI, FARID
; APPLICANT: NEDJARI, HAFED
; APPLICANT: GLASER, PHILIPPE
; APPLICANT: KUNST, FRANCK
; APPLICANT: COSSART, PASCALE
; APPLICANT: DANIELS, JUSTIN
; APPLICANT: GOEBEL, WERNER
; APPLICANT: KREFT, JURGEN
; APPLICANT: KUHN, MICHAEL
; APPLICANT: NG, EVA
; APPLICANT: VAZQUEZ-BOLAND, ANTONIO
; APPLICANT: DOMINGUEZ-BERNAL, GUSTAVO
; APPLICANT: GARRIDO-GARCIA, PATRICIA
; APPLICANT: TIERREZ-MARTINEZ, ALBERTO
; APPLICANT: AMEND, ALEXANDRA
; APPLICANT: CHAKRABORTY, TRINAD
; APPLICANT: DOMANN, EUGEN
; APPLICANT: HAIN, THORSTEN
; APPLICANT: BERCHE, PATRICK
; APPLICANT: CHARBIT, ALAIN
; APPLICANT: DURANT, LIONEL
; APPLICANT: PEREZ-DIAZ, JOSE-CLAUDIO
```

```

; APPLICANT: BAQUERO, FERNANDO
; APPLICANT: GARCIA DEL PORTILLO, FRANCISCO
; APPLICANT: GOMEZ-LOPEZ, NURIA
; APPLICANT: MADUENIO, ENCARNIA
; APPLICANT: PABLOS, BETRIZ DE
; APPLICANT: WEHLAND, JURGEN
; APPLICANT: KARST, UWE
; APPLICANT: ENTIAN, KARL-DIETER
; APPLICANT: HAUF, JORG
; APPLICANT: ROSE, MATTHIAS
; APPLICANT: VOSS, HAMUT
; TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
; FILE REFERENCE: 05394.0018-02
; CURRENT APPLICATION NUMBER: US/11/045,004
; CURRENT FILING DATE: 2005-01-28
; PRIOR APPLICATION NUMBER: 10/637,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: 10/257,023
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: PCT/FR01/01118
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: FR 00/04,629
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 2854
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1742
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-11-045-004-1742

Query Match          8.7%; Score 74; DB 7; Length 357;
Best Local Similarity 26.6%; Pred. No. 0.91;
Matches 25; Conservative 18; Mismatches 33; Indels 18; Gaps 4;

Qy 29 KENPTIDLENSYVP-----LDNQYKGIKRVDEHPEYVLDHDKVWAT--ATYNNDLNGIKTNDIM 82
Db 99 EENLLIPLDHSKLPNEKYLDPRFMDLSFDDDNKYSMPYFWGTGLIYINKEMFPDKNFDTW 158

Qy 83 LGVVIPTD-----BEDVGLGM-ELGYALS 104
Db 159 NALFDPFLKNQIILLIDGAREVMGLGNSLGLYSLN 192

RESULT 4
US-10-556-060-326
; Sequence 326, Application US/10556060
; GENERAL INFORMATION:
; APPLICANT: MEINKS, ANDREAS
; APPLICANT: NAGY, ESZTER
; APPLICANT: HANNER, MARKUS
; APPLICANT: HORKY, MARKUS
; APPLICANT: KALLEDA, SABINE
; APPLICANT: PRUSTOMERSKY, SONJA
; TITLE OF INVENTION: S. AGALACTIAE ANTIGENS I + II
; FILE REFERENCE: SONN:080US
; CURRENT APPLICATION NUMBER: US/10/556,060
; CURRENT FILING DATE: 2005-11-07
; PRIOR APPLICATION NUMBER: PCT/EP2004/004856
; PRIOR FILING DATE: 2004-05-06
; PRIOR APPLICATION NUMBER: 03450112.2
; PRIOR FILING DATE: 2003-05-07
; PRIOR APPLICATION NUMBER: 03450266.6
; PRIOR FILING DATE: 2003-11-28
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 326
; LENGTH: 1310
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-556-060-326

Query Match          8.6%; Score 73; DB 6; Length 1310;
```



[illegible]

```

RESULT 5
US-60-732-162-404
; Sequence 404, Application US/60732162
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Debrus, Sophie
; APPLICANT: Berdewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated with Asthma Disease
; FILE REFERENCE: 059908-5010-PR
; CURRENT APPLICATION NUMBER: US/60/732,162
; CURRENT FILING DATE: 2005-11-02
; NUMBER OF SEQ ID NOS: 4417
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 404
; LENGTH: 578
; TYPE: PR
; ORGANISM: Homosapiens
US-60-732-162-404

```

Query Match	8.5%;	Score 72.5;	DB 8;	Length 578;
Best Local Similarity	23.9%;	Pred. No. 2.6;		
Matches	39;	Conservative 20;	Mismatches 51;	Indels 53; Gaps 8;
Qy	3	KTITYFGAGWFTDRQKAY-----KEAMEALKENPTIDL-----ENSYV 41		
Db	386	KKYVWFILIGWYADNWFKIYDPSINCTVDEMTAEVGEHITTEIVMLNPANTESISNMTSQE 445		
Qy	42	PLDQYKGIIRVDEHPE-----YLHKDWATATYNNDLGKLTINDIMLGVIYPODEE 92		
Db	446	FVEKLTWK--RLKRPEETGGFGEAPLAYDAITWALALAIN-----KTS-----485		
Qy	93	VGLGMEIGALYSQGYVLLVTPDEDYKGKPINLMSW-GVSDNVI 134		
Db	486	-GGGSRGSRVRLEDFYNNNOTITTDQIY-RAMSSSPFEGYSGHVV 526		

RESULT 6  
US-60-742-219-464  
; Sequence 464, Application US/60742219  
; GENERAL INFORMATION:  
; APPLICANT: Belouchi, Abdelmajid  
; APPLICANT: Raelson, John Verner  
; APPLICANT: Bradley, Walter Edward  
; APPLICANT: Paquin, Bruno  
; APPLICANT: Nguyen-Huu, Quynh  
; APPLICANT: Croteau, Pascal  
; APPLICANT: Allard, Rene  
; APPLICANT: Little, Randall David  
; APPLICANT: Keith, Tim  
; APPLICANT: Cousineau, Johanne  
; APPLICANT: Eerdewegh, Paul Van



Qy 93 VGLGMEGLGVALSQGKYVLLVLPDEDYDKPILMSW-GVSDNVI 134  
Db 369 -GGGSGVRLEDFNYYNNQTTDQIY-RAMNSSSFEVSGHV 409

## RESULT 11

US-60-742-219-462  
; Sequence 462, Application US/60742219  
; GENERAL INFORMATION:  
; APPLICANT: Belouchi, Abdelmajid  
; APPLICANT: Raelson, John Verner  
; APPLICANT: Bradley, Walter Edward  
; APPLICANT: Paquin, Bruno  
; APPLICANT: Nguyen-Huu, Quynh  
; APPLICANT: Croteau, Pascal  
; APPLICANT: Allard, Rene  
; APPLICANT: Little, Randall David  
; APPLICANT: Keith, Tim  
; APPLICANT: Cousineau, Johanne  
; APPLICANT: Eerdewegh, Paul Van  
; APPLICANT: Segal, Jonathan  
; TITLE OF INVENTION: Gene Map of the Human Genes Associated with Psoriasis  
; FILE REFERENCE: 306522-2000  
; CURRENT APPLICATION NUMBER: US/60/742,219  
; CURRENT FILING DATE: 2005-12-05  
; NUMBER OF SEQ ID NOS: 7303  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 462  
; LENGTH: 844  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-60-742-219-462

Query Match 8.3%; Score 70.5; DB 8; Length 844;  
Best Local Similarity 23.9%; Pred. No. 7.1;  
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;  
Qy 3 KKTIFYGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----ENSYV 41  
Db 269 KKYVWFLIGWYADNWFKYDPSINCTVDMEAVEGHITTEIVMLNPANTRISISNMTSQE 328  
Qy 42 PLDNOYKGRVDEHPE-----YLDKQWATATYNDLNGIKTNDIMLGVIYPDEED 92  
Db 329 FVEKLTK--RLKRHPETGGFQEAFLAYDAIWAALALN-----KTS----- 368  
Qy 93 VGLGMEGLGVALSQGKYVLLVLPDEDYDKPILMSW-GVSDNVI 134  
Db 369 -GGGSGVRLEDFNYYNNQTTDQIY-RAMNSSSFEVSGHV 409

## RESULT 12

US-60-732-162-408  
; Sequence 408, Application US/60732162  
; GENERAL INFORMATION:  
; APPLICANT: Belouchi, Abdelmajid  
; APPLICANT: Raelson, John V  
; APPLICANT: Bradley, Walter E  
; APPLICANT: Paquin, Bruno  
; APPLICANT: Fournier, Helene  
; APPLICANT: Nguyen-Huu, Quynh  
; APPLICANT: Croteau, Pascal  
; APPLICANT: Allard, Rene  
; APPLICANT: Debrus, Sophie  
; APPLICANT: Eerdewegh, Paul V  
; APPLICANT: Little, Randall D  
; APPLICANT: Keith, Tim  
; APPLICANT: Segal, Jonathan  
; TITLE OF INVENTION: Genemap of the Human Genes Associated With Asthma Disease  
; FILE REFERENCE: 059908-5010-PR  
; CURRENT APPLICATION NUMBER: US/60/732,162  
; CURRENT FILING DATE: 2005-11-02  
; NUMBER OF SEQ ID NOS: 4417

; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 408  
; LENGTH: 899  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-60-732-162-408

Query Match 8.3%; Score 70.5; DB 8; Length 899;  
Best Local Similarity 23.9%; Pred. No. 7.8;  
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;  
Qy 3 KKTIFYGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----ENSYV 41  
Db 324 KKYVWFLIGWYADNWFKYDPSINCTVDMEAVEGHITTEIVMLNPANTRISISNMTSQE 383  
Qy 42 PLDNOYKGRVDEHPE-----YLDKQWATATYNDLNGIKTNDIMLGVIYPDEED 92  
Db 384 FVEKLTK--RLKRHPETGGFQEAFLAYDAIWAALALN-----KTS----- 423  
Qy 93 VGLGMEGLGVALSQGKYVLLVLPDEDYDKPILMSW-GVSDNVI 134  
Db 424 -GGGSGVRLEDFNYYNNQTTDQIY-RAMNSSSFEVSGHV 464

## RESULT 13

US-60-742-219-466  
; Sequence 466, Application US/60742219  
; GENERAL INFORMATION:  
; APPLICANT: Belouchi, Abdelmajid  
; APPLICANT: Raelson, John Verner  
; APPLICANT: Bradley, Walter Edward  
; APPLICANT: Paquin, Bruno  
; APPLICANT: Nguyen-Huu, Quynh  
; APPLICANT: Croteau, Pascal  
; APPLICANT: Allard, Rene  
; APPLICANT: Little, Randall David  
; APPLICANT: Keith, Tim  
; APPLICANT: Cousineau, Johanne  
; APPLICANT: Eerdewegh, Paul Van  
; APPLICANT: Segal, Jonathan  
; TITLE OF INVENTION: Gene Map of the Human Genes Associated with Psoriasis  
; FILE REFERENCE: 306522-2000  
; CURRENT APPLICATION NUMBER: US/60/742,219  
; CURRENT FILING DATE: 2005-12-05  
; NUMBER OF SEQ ID NOS: 7303  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 466  
; LENGTH: 899  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-60-742-219-466

Query Match 8.3%; Score 70.5; DB 8; Length 899;  
Best Local Similarity 23.9%; Pred. No. 7.8;  
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;  
Qy 3 KKTIFYGAGWFTDRQNKAY-----KEAMEALKENPTIDL-----ENSYV 41  
Db 324 KKYVWFLIGWYADNWFKYDPSINCTVDMEAVEGHITTEIVMLNPANTRISISNMTSQE 383  
Qy 42 PLDNOYKGRVDEHPE-----YLDKQWATATYNDLNGIKTNDIMLGVIYPDEED 92  
Db 384 FVEKLTK--RLKRHPETGGFQEAFLAYDAIWAALALN-----KTS----- 423  
Qy 93 VGLGMEGLGVALSQGKYVLLVLPDEDYDKPILMSW-GVSDNVI 134  
Db 424 -GGGSGVRLEDFNYYNNQTTDQIY-RAMNSSSFEVSGHV 464

## RESULT 14

US-60-732-162-406  
; Sequence 406, Application US/60732162  
; GENERAL INFORMATION:

; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John V
; APPLICANT: Bradley, Walter E
; APPLICANT: Paquin, Bruno
; APPLICANT: Fournier, Helene
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Debrus, Sophie
; APPLICANT: Eerdewegh, Paul V
; APPLICANT: Little, Randall D
; APPLICANT: Keith, Tim
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Genemap of the Human Genes Associated With Asthma Disease
; FILE REFERENCE: 059908-5010-PR
; CURRENT APPLICATION NUMBER: US/60/732,162
; CURRENT FILING DATE: 2005-11-02
; NUMBER OF SEQ ID NOS: 4417
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 406
; LENGTH: 961
; TYPE: PR
; ORGANISM: Homosapiens
US-60-732-162-406

Query Match 8.3%; Score 70.5; DB 8; Length 961;
Best Local Similarity 23.9%; Pred. No. 8.5;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;

Qy 3 KKTIFYGAGWFTDRONKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db 386 KKYVWFLIGWYADNWFKIYDPSINCTVDTEAVEGHITTEIVMLNPANTRISINMTSQE 445
Qy 42 PLDNQYKGIKRVDEHPE-----YLHDKVWATATYNNDLNGIKTNDIMLGYYIPDEED 92
Db 446 FVEKLTGK--RLKRHPETGGFQEPALYDAIWALALN-----KTS----- 485
Qy 93 VGLGMELGYALSQKGYVLLVLPDSDYKPKINLMSW-GVSDNVI 134
Db 486 -GGGGRSGVRLEDFNYYNQITTDQIY-RAMNSSSFEGVSGHV 526

RESULT 15
US-60-742-219-460
; Sequence 460, Application US/60/42219
; GENERAL INFORMATION:
; APPLICANT: Belouchi, Abdelmajid
; APPLICANT: Raelson, John Verner
; APPLICANT: Bradley, Walter Edward
; APPLICANT: Paquin, Bruno
; APPLICANT: Nguyen-Huu, Quynh
; APPLICANT: Croteau, Pascal
; APPLICANT: Allard, Rene
; APPLICANT: Little, Randall David
; APPLICANT: Keith, Tim
; APPLICANT: Cousineau, Johanne
; APPLICANT: Eerdewegh, Paul Van
; APPLICANT: Segal, Jonathan
; TITLE OF INVENTION: Gene Map of the Human Genes Associated with Psoriasis
; FILE REFERENCE: 306522-2000
; CURRENT APPLICATION NUMBER: US/60/742,219
; CURRENT FILING DATE: 2005-12-05
; NUMBER OF SEQ ID NOS: 7303
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 460
; LENGTH: 961
; TYPE: PR
; ORGANISM: Homo sapiens
US-60-742-219-460

Query Match 8.3%; Score 70.5; DB 8; Length 961;
Best Local Similarity 23.9%; Pred. No. 8.5;
Matches 39; Conservative 20; Mismatches 51; Indels 53; Gaps 8;

Qy 3 KKTIFYGAGWFTDRONKAY-----KEAMEALKENPTIDL-----ENSYV 41
Db 386 KKYVWFLIGWYADNWFKIYDPSINCTVDTEAVEGHITTEIVMLNPANTRISINMTSQE 445
Qy 42 PLDNQYKGIKRVDEHPE-----YLHDKVWATATYNNDLNGIKTNDIMLGYYIPDEED 92
Db 446 FVEKLTGK--RLKRHPETGGFQEPALYDAIWALALN-----KTS----- 485
Qy 93 VGLGMELGYALSQKGYVLLVLPDSDYKPKINLMSW-GVSDNVI 134
Db 486 -GGGGRSGVRLEDFNYYNQITTDQIY-RAMNSSSFEGVSGHV 526

Search completed: December 23, 2005, 22:17:35
Job time : 15 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:50:25 ; Search time 38 Seconds  
(without alignments)  
397.527 Million cell updates/sec

Title: US-10-049-750-14  
Perfect score: 848  
Sequence: 1 MPKTIYFGAGWFTDRQNK.....QLKDFNFKPRDFYEGAVY 157

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	637	75.1	158	JC7522	nucleoside deoxyri
2	136.5	16.1	159	E86686	hypothetical prote
3	91	10.7	719	A42808	IG light chain-bin
4	90.5	10.7	660	G90330	amine oxidase (cop
5	86.5	10.2	462	A86717	conserved hypothet
6	83	9.8	206	B28439	maturase-related h
7	83	9.8	369	F84113	hypothetical prote
8	82.5	9.7	376	G71925	cag island protein
9	82.5	9.7	766	A40258	RAS GTPase-activat
10	82	9.7	1088	S39261	VPI protein - porc
11	82	9.7	1838	A42091	transcription acti
12	81.5	9.6	376	A64587	cag pathogenicity
13	81.5	9.6	2172	T00936	probable ATP-depen
14	81	9.6	241	H95078	hypothetical prote
15	81	9.6	313	S47433	cathepsin L (EC 3.
16	80.5	9.5	2233	S63347	acetyl-CoA carboxy
17	80	9.4	303	T17774	hypothetical prote
18	79.5	9.4	446	H97091	chitinase family p
19	79.5	9.4	1452	A97323	DNA polymerase III
20	78.5	9.3	236	H85435	hypothetical prote
21	78.5	9.3	423	C69936	conserved hypothet
22	78	9.2	410	C84827	hypothetical prote
23	78	9.2	751	HYRBN	neprilysin (EC 3.4
24	78	9.2	816	A71006	hypothetical prote
25	77.5	9.1	816	C83917	DNA topoisomerase
26	77.5	9.1	2216	S78398	hypothetical prote
27	77	9.1	1088	P1XRBR	inner layer protei
28	77	9.1	1088	S13558	VPI protein - bovi
29	77	9.1	1650	T18444	hypothetical prote

30 76.5 9.0 241 2 E64476 hypothetical prote  
31 76.5 9.0 432 2 G71695 proline-tRNA ligas  
32 76 9.0 435 2 H75443 aspartyl-tRNA synt  
33 76 9.0 455 2 A45063 immunoglobulin-bin  
34 76 9.0 519 2 S45723 p60 protein - oat  
35 76 9.0 2269 2 T28677 rhothy protein -  
36 75.5 8.9 177 1 KIVZSW thymidine kinase (  
37 75.5 8.9 181 1 KIVZSK hypothetrical prote  
38 75.5 8.9 256 2 D89789 hypothetrical prote  
39 75.5 8.9 437 2 B98074 probable TonB-depe  
40 75.5 8.9 690 2 AB0124 hypothetrical prote  
41 75.5 8.9 759 2 T00153 hypothetrical prote  
42 75.5 8.9 1177 2 I64233 hypothetrical prote  
43 75.5 8.9 1260 2 H89984 hypothetrical prote  
44 75.5 8.9 1504 2 T17426 FK506 polyketide s  
45 75 8.8 602 2 E64464 hypothetrical prote

ALIGNMENTS

RESULT 1

JC7522  
nucleoside deoxyribosyltransferase (EC 2.4.2.6) II - Lactobacillus helveticus  
C:Species: Lactobacillus helveticus  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 09-Jul-2004  
C:Accession: JC7522; PC7103  
R:Okuyama, K.; Noguchi, T.  
Biosci. Biotechnol. Biochem. 64, 2243-2245, 2000  
A:Title: Molecular cloning and expression of the nucleoside deoxyribosyltransferase-II g  
A:Reference number: JC7522; MUID:21012342; PMID:11129605  
A:Accession: JC7522  
A:Molecule type: DNA  
A:Residues: 1-158 <OKU>  
A:Cross-references: UNIPROT:Q9KWF0; UNIPARC:UPI000015C9E3; DDBJ:AB039914  
A:Experimental source: strain ATCC 8018  
A:Accession: PC7103  
A:Molecule type: protein  
A:Residues: 1-11;53-62 <OK2>  
A:Cross-references: UNIPARC:UPI000017A510; UNIPARC:UPI000017A511  
A:Comment: This enzyme catalyzes transfer of glycosyl residues from a donor deoxynucleos  
C:Genetics:  
A:Gene: ndtB  
C:Keywords: glycosyltransferase; pentosyltransferase

Query Match	75.1%	Score 637;	DB 2;	Length 158;
Best Local Similarity	73.5%;	Pred. No. 8.9e-49;		
Matches 114;	Conservative 21;	Mismatches 20;	Indels 0;	Gaps 0;
QY	3	KKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSVPLDNQYKGRVDEHPEYLHDK	62	
Db	4	KKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSVPLDNQYKGRVDEHPEYLHDK	63	
QY	63	VWATATYNDLNGIKTNDIMLGVIYIPDEEDVGLGELGVALSQGVLLVIFDEYDGKPI	122	
Db	64	EWASATYNDLNGIKTNDIMLGVIYIPDEEDVGLGELGVALSQGVLLVIFDEYDGKPI	123	
QY	123	NLMMSGVSDNVIKMSQLKDFNFKPRDFYEGAVY	157	
Db	124	ILMSGVCDNASQISELKDFDFNFKRYNFDGAVY	158	

RESULT 2

E86686  
hypothetical protein vejD [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: E86686  
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: E86686









submitted to the EMBL Data Library, December 1997  
A;Description: Arabidopsis thaliana chromosome II BAC T24P15 genomic sequence.  
A;Reference number: Z14212  
A;Accession: T00936  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2172 <ROU>  
A;Cross-references: UNIPROT:O48534; UNIPARC:UPI00000AB148; EMBL:AC002561; NID:g2673901;  
A;Experimental source: cultivar Columbia  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: H84851  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1621, 'M', 1623-2172 <STO>  
A;Cross-references: UNIPARC:UPI000017AF4A; GB:AE002093; NID:g2673917; PIDN:AA888651.1; G  
C;Genetics:  
A;Gene: T24P15.18; At2g42270  
A;Map position: 2

Query Match 9.6%; Score 81.5; DB 2; Length 2172;  
Best Local Similarity 23.2%; Pred. No. 1.3e+02;  
Matches 39; Conservative 27; Mismatches 61; Indels 41; Gaps 8;  
Qy 1 MPKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVP-----LDNQYKGRVDEH 55  
Db 435 MDLESAPNQGGFTRENNKC-----ELPDRFRIRGKEFDEHVHPWVSKFDSNEKLKVTISDL 492  
Qy 56 PEY-----LHDKVWATATYND-----LNGIKTNDIMLGVTIPDEEDVGL 95  
Db 493 PEWAQPAFRGQQLNRVQSVGYGTALFKADNILLCAPTGAGTINVAVTIL-----HQGL 548  
Qy 96 GMEIGYALSQGYVLLVTPDEYDYGKPINLMSWGVSDNVKMSQ-LKDF 142  
Db 549 NNNPGGTGNGYKIV-----YVAPMKAL---VAEVVDSLSQRLKDF 587

RESULT 14  
H95078  
hypothetical protein SP0680 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C;Species: Streptococcus pneumoniae  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C;Accession: H95078  
R;Fetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eissen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A;Reference number: A95000; MUID:21357209; PMID:11463916  
A;Accession: H95078  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-241 <KUR>  
A;Cross-references: UNIPROT:Q97RV6; UNIPARC:UPI000005153B; GB:AE005672; PIDN:AAK74825.1;  
A;Experimental source: strain TIGR4  
C;Genetics:  
A;Gene: SP0680  
C;Superfamily: conserved hypothetical protein H11243

Query Match 9.6%; Score 81; DB 2; Length 241;  
Best Local Similarity 24.9%; Pred. No. 8.5;  
Matches 47; Conservative 19; Mismatches 59; Indels 64; Gaps 10;  
Qy 13 FTDRONKAYKEAMEAL-----KENPTI-DLENSYVPLDNQYKGRVDE----- 54  
Db 49 FODRIIQGYEHTYLMHLHPAGAVTANKKELPTVMDLLPSNIQSKLYAVGRLDRDTGL 108  
Qy 55 -----HPEYLHDKVWATATYNDLNGIKTND-----IMLGVIYPDE----- 90

Db 109 LLLTDNGLQFGLLHPQYHVDK-----TYQEVNGLLTPDHIQTFQKGIIVLDDTVCKPA 163  
Qy 91 --EDVGLGMEIGYA---LSQGYVLLVTPDEYDYGKPINLMSWGVSDNVKMSQLKDFNFN 145  
Db 164 KLEILSASPSLSQASITISEGKPHQI-----KKM-FLSVGVKVTSLKRIQFGDFTLN 214  
Qy 146 KPRDFYFEG 154  
Db 215 P---DLAEG 220  
RESULT 15  
S47433  
cathepsin L (EC 3.4.22.15) - Norway lobster  
C;Species: Nephrops norvegicus (Norway lobster)  
C;Date: 23-Nov-1994 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C;Accession: S47433  
R;le Boulay, C.; van Wormhoudt, A.; Sellos, D.  
submitted to the EMBL Data Library, August 1994  
A;Description: Molecular cloning and sequencing of the two cDNAs that encode cathepsin L  
A;Reference number: S47432  
A;Accession: S47433  
A;Molecule type: mRNA  
A;Residues: 1-313 <LEB>  
A;Cross-references: UNIPROT:Q27708; UNIPARC:UPI0000084019; EMBL:X80990; NID:g530735; PID  
C;Superfamily: papain  
C;Keywords: cysteine proteinase; hydrolase  
F;122,260,280/Active site: Cys, His, Asn #status predicted

Query Match 9.6%; Score 81; DB 2; Length 313;  
Best Local Similarity 26.3%; Pred. No. 12;  
Matches 44; Conservative 15; Mismatches 48; Indels 60; Gaps 11;  
Qy 9 GAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNQYKGRVDP-----E 54  
Db 163 GGGWMT-----SAFDYIKDNGIDTSSY-PYEAQDRSCRFDANSIGATCTGFEVQ 213  
Qy 55 H-PYLLHDKVWATATYNDLNGIKT-----NDIMLGVI-----PDEEDVGLGMEL 99  
Db 214 HTEERALHEAV-----SDIGPISVAIDASHFSPQFYSSGVYKCKSPTNLDHGV-LAV 265  
Qy 100 GYALSQGYVLLVTPDEYDYGKPINLMSWGV---SDNVKMSQLKDFN 143  
Db 266 GYGTES-----TEDYWLVKN--SWGSGMGDAGYIKMSRNRDNN 301

Search completed: December 23, 2005, 22:04:55  
Job time : 42 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2005, 21:48:35 ; Search time 230 Seconds  
(without alignments)  
481.600 Million cell updates/sec

Title: US-10-049-750-14  
Perfect score: 848  
Sequence: 1 MPKTYIFGAGWFTDRONKA.....QLKDFNFKNRPFDFYEGAVY 157

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	843	99.4	156	1 NTD_LACLE	Q9r5v5 lactobacill
2	728	85.8	158	2 Q8RLY4 LACHE	Q8rlY4 lactobacill
3	719	84.8	159	2 Q5FMH8 LACAC	Q5fmH8 lactobacill
4	653	77.0	149	1 NTD_LACJO	Q74lq9 lactobacill
5	634	74.8	158	2 Q9KWF0 LACHE	Q9kwF0 lactobacill
6	209	24.6	168	1 NTD_LACFE	Q6yni5 lactobacill
7	171.5	20.2	167	2 Q8RLY5 LACHE	Q8rlY5 lactobacill
8	154.5	18.2	167	2 Q5FIN0 LACAC	Q5fin0 lactobacill
9	136.5	16.1	159	2 Q9CI73 LACLA	Q9ci73 lactococcus
10	124.5	14.7	157	2 Q88SL4 LACPL	Q88sl4 lactococcus
11	106	12.5	155	2 Q4QAY0 LEIMA	Q4qay0 leishmania
12	96	11.3	146	2 Q88X23 LACPL	Q88x23 lactobacill
13	96	11.3	1149	2 Q5EMR0 AMEPV	Q5emr0 amacta moo
14	95	11.2	1029	2 Q5ZMQ1 CHICK	Q5zmql gallus gall
15	93.5	11.0	264	2 Q420Q6 PLABE	Q420q6 plasmodium
16	93.5	11.0	325	2 Q7P6V5 FUSNV	Q7p6v5 fusobacteri
17	92.5	10.9	800	2 Q4L5E4 STAHJ	Q4l5e4 staphylococ
18	92	10.8	553	1 ADSC METAC	Q8trh5 methanosarc
19	91	10.7	719	2 Q51912 PEPMA	Q51912 peptostrept
20	90.5	10.7	660	2 Q97XM1 SULSO	Q97xm1 sulfolobus
21	90	10.6	454	2 Q4JUB7 CORJK	Q4jub7 corynebacte
22	89	10.5	218	2 Q8RHH4 FUSNV	Q8rhh4 fusobacteri
23	89	10.5	462	2 Q7WST1 PSPPA	Q7wst1 pseudomonas
24	89	10.5	844	2 Q7R989 PLAYO	Q7r989 plasmodium
25	89	10.5	1139	2 Q7RRP6 PLAYO	Q7rxp6 plasmodium
26	88	10.4	504	1 AMPA CANBF	Q7vqt0 candidatus
27	88	10.4	544	2 Q7P5F2 FUSNV	Q7p5f2 fusobacteri
28	87.5	10.3	519	2 Q7RT73 PLAYO	Q7rt73 plasmodium
29	87	10.3	678	2 Q8TMR7 METAC	Q8tmr7 methanosarc
30	86.5	10.2	462	2 Q9CHJ4 LACIA	Q9chj4 lactococcus
31	85.5	10.1	438	2 Q9H804 HUMAN	Q9h804 homo sapien

32	85.5	10.1	487	2	Q9NZP8 HUMAN	Q9nZp8 homo sapien
33	85	10.0	446	1	TIG_EHRRG	Q5ffg8 ehrlichia r
34	85	10.0	480	2	Q7MQX8 WOLSU	Q7mqX8 wollinella s
35	85	10.0	606	1	BTUB_PROPR	Q6llj3 photobacter
36	85	10.0	725	2	Q91YT4 MOUSE	Q9lyt4 mus musculu
37	85	10.0	993	2	Q8BTX6 MOUSE	Q8btX6 mus musculu
38	85	10.0	1040	1	SK2L2 MOUSE	Q9czu3 mus musculu
39	85	10.0	1568	2	Q6P9P2 BRARE	Q6p9P2 brachydanio
40	84.5	10.0	154	2	Q7VB88 PROMA	Q7vB88 prochloroco
41	84.5	10.0	2985	2	Q4UAS6 THEAN	Q4uas6 theileria a
42	84.5	10.0	3091	2	Q4UD69 THEAN	Q4ud69 theileria a
43	84	9.9	351	2	Q89SM8 CLOTE	Q89sm8 clostridium
44	84	9.9	446	1	TIG_EHRRW	Q5hbX6 ehrlichia r
45	84	9.9	596	2	Q8N5R0 HUMAN	Q8n5r0 homo sapien

## ALIGNMENTS

### RESULT 1

NTD\_LACLE STANDARD; PRT; 156 AA.  
AC Q9R5V5;

DT 25-OCT-2004 (Rel. 45, Created)

DT 25-OCT-2004 (Rel. 45, Last sequence update)

DT 01-FEB-2005 (Rel. 46, Last annotation update)

DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-deoxyribosyltransferase).

GN Name=ntd;

OS Lactobacillus leichmannii.

OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;

OC Lactobacillus.

OX NCBI\_TaxID=28039;

RN [1]

RP NUCLEOTIDE SEQUENCE, PROTEIN SEQUENCE OF 1-25, FUNCTION, ACTIVE SITE, AND MUTAGENESIS OF GLU-97.

RX MEDLINE=95318137; PubMed=7797550; DOI=10.1074/jbc.270.26.15551;

RA Porter D.J.T., Merrill B.M., Short S.A.;

RT "Identification of the active site nucleophile in nucleoside 2-deoxyribosyltransferase as glutamic acid 98.";

RL J. Biol. Chem. 270:15551-15556(1995).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).

RX MEDLINE=96419140; PubMed=8805514;

RA Armstrong S.R., Cook W.J., Short S.A., Balick S.E.;

RT "Crystal structures of nucleoside 2-deoxyribosyltransferase in native and ligand-bound forms reveal architecture of the active site.";

RL Structure 4:97-107(1996).

CC -!- FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-deoxyribonucleosides and the transfer of the deoxyribosyl moiety to an acceptor purine or pyrimidine base.

CC -!- CATALYTIC ACTIVITY: 2-deoxy-D-ribose(1) + base(2) = 2-deoxy-D-ribose(1)-base(2) + base(1).

CC -!- BIOPHYSICOCHEMICAL PROPERTIES:

CC pH dependence:

CC Optimum pH is 6.0;

CC -!- PATHWAY: Nucleotide metabolism.

CC -!- SUBUNIT: Homohexamer.

CC -!- SIMILARITY: Belongs to the nucleoside deoxyribosyltransferase family.

CC -!- CAUTION: Was originally (Ref.1) thought to originate from E.coli.

-----  
This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.  
-----  
PDB: 1F8X; X-ray; A/B=1-156.  
PDB: 1F8Y; X-ray; A/B=1-156.  
InterPro: IPR007710; N deoxyrib trans.  
Pfam: PF05014; Nuc\_deoxyrib\_tr; 1.  
3D-structure; Direct protein sequencing; Nucleotide metabolism;  
-----

```
KW Transferase.
FT INIT_MET 0 0 Nucleophile.
FT CITE 97 97 E->A: Loss of transferase activity.
FT MUTAGEN 97 97
SQ SEQUENCE 156 AA; 17949 MW; 3A3ABC3FD5B4743B CRC64;

Query Match 99.4%; Score 843; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. NO. 4.6e-64;
Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKRVDEHPEYLHD 61
DB 1 PKTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKRVDEHPEYLHD 60
QY 62 KWATATYNNDLNGIKTNDIMLGVIYIPDEEDVGLGMELGYALSQKGYVLLVIPDEDYGKP 121
DB 61 KWATATYNNDLNGIKTNDIMLGVIYIPDEEDVGLGMELGYALSQKGYVLLVIPDEDYGKP 120
QY 122 INLMSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 121 INLMSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 156

RESULT 2
Q8RLY4_LACHE PRELIMINARY; PRT; 158 AA.
AC Q8RLY4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 24, Last annotation update)
DE N-deoxyribosyltransferase.
GN Name=ntd;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CNR232;
RX MEDLINE=21964041; PubMed=11836245; DOI=10.1074/jbc.M111995200;
RA Kaminski P.A.;
RT "Functional cloning, heterologous expression, and purification of two different N-deoxyribosyltransferases from Lactobacillus helveticus.";
RL J. Biol. Chem. 277:14400-14407(2002).
DR EMBL; AY064167; AAL73114.1; -; Genomic_DNA.
DR HSP; Q9RSV5; 1F8Y.
DR SMR; Q8RLY4; 4-158.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Transferase.
SQ SEQUENCE 158 AA; 18148 MW; 00FC9E4B210C47B8 CRC64;

Query Match 85.8%; Score 728; DB 2; Length 158;
Best Local Similarity 82.6%; Pred. No. 3.1e-54;
Matches 128; Conservative 20; Mismatches 7; Indels 0; Gaps 0;

QY 3 KKTIFYGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKRVDEHPEYLHDK 62
DB 4 KKTIFYGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLENENYKGIKRVDEHPEYLHNI 63
QY 63 VWATATYNNDLNGIKTNDIMLGVIYIPDEEDVGLGMELGYALSQKGYVLLVIPDEDYGKPI 122
DB 64 EWASATYHNDLVGIKTSDVMGLVYLPEEDVGLGMELGYALSQKGYVLLVIPDEDYGKPI 123
QY 123 NLSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 124 NLSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 158

RESULT 3
Q5FMM8_LACAC PRELIMINARY; PRT; 159 AA.
ID Q5FMM8_LACAC
```

```
Q5FMM8;
AC Q5FMM8;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE 2-deoxyribosyltransferase.
GN OrderedLocusNames=LBA0145;
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RX PubMed=15671160; DOI=10.1073/pnas.0409188102;
RA Altermann E., Russell W.M., Azcarate-Peril M.A., Barrangou R., Buck B.L., McAuliffe O., Souther N., Dobson A., Duong T., Callanan M., Lick S., Hanrick A., Cano R., Klaenhammer T.R.;
RA "Complete genome sequence of the probiotic lactic acid bacterium Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912(2005).
DR EMBL; CP000033; AAV42046.1; -; Genomic DNA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 159 AA; 18252 MW; 5CC7FFDBA57938FE CRC64;

Query Match 84.8%; Score 719; DB 2; Length 159;
Best Local Similarity 83.1%; Pred. No. 1.8e-53;
Matches 128; Conservative 18; Mismatches 8; Indels 0; Gaps 0;

QY 4 KTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKRVDEHPEYLHDKV 63
DB 6 KTIYFGAGWFTDRONKAYKEAMEALKENPTIDLENSYVPLENENYKGIKRVDEHPEYLHDIE 65
QY 64 WATATYNNDLNGIKTNDIMLGVIYIPDEEDVGLGMELGYALSQKGYVLLVIPDEDYGKPIN 123
DB 66 WASATYHNDLVGIKTSDIMLGVIYIPDEEDVGLGMELGYALSQKGYVLLVIPDEDYGKPIN 125
QY 124 LMSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 157
DB 126 LMSWGVSDNVIKMSQLKDFNFKPRDFYEGAVY 159

RESULT 4
NTD_LACJO
ID NTD_LACJO STANDARD; PRT; 149 AA.
AC Q74LQ9;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 13-SRP-2005 (Rel. 48, Last annotation update)
DE Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-deoxyribosyltransferase).
GN Name=ntd; OrderedLocusNames=LJ0124;
OS Lactobacillus johnsonii.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=33959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=NCC 533;
RX PubMed=14983040; DOI=10.1073/pnas.0307327101;
RA Priddy R.D., Berger B., Desiere F., Vilanova D., Barretto C., Pittet A.-C., Zwahlen M.-C., Rouvet M., Altermann E., Barrangou R., Mollet B., Mercenier A., Klaenhammer T., Arigoni F., Schell M.A.;
RT "The genome sequence of the probiotic intestinal bacterium Lactobacillus johnsonii NCC 533.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:2512-2517(2004).
CC -I- FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-deoxyribonucleosides and the transfer of the deoxyribosyl moiety to an acceptor purine or pyrimidine base (By similarity).
CC -I- CATALYTIC ACTIVITY: 2-deoxy-D-ribose(1) + base(2) = 2-deoxy-D-ribose(1)-base(2) + base(1).
```

QY	63	WVATATYNDNLGIKTNIMLGVYIPEEDVDGLGMELGYALSOQKYVLLVIPDEDYGKPI	123
Db	64	EWASATYHNDLGIKTSVDVLLGVYLFQEEHVGGLGMELGYPLSOGLFFWFESHMKDYGKPI	123
QY	123	NLMSWGVSNDNVIRKMSQLKDFNFKNPRFDFYEGAVY	157
Db	124	ILMSWGVCDNASQISELKDFDFNKPKNRYFYDGAVY	158

NTD_LACFE	NTD_LACPE	STANDARD;	PRT;	168	AA.
AC	Q5YN15;				
DT	25-OCT-2004	(Rel. 45, Created)			
DT	25-OCT-2004	(Rel. 45, Last sequence update)			
DT	13-SEP-2005	(Rel. 48, Last annotation update)			
DE	Nucleoside deoxyribosyltransferase (EC 2.4.2.6) (N-deoxyribosyltransferase).				
DE	Name:ntdi;				
GN	Lactobacillus fermentum.				
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;				
OC	Lactobacillus.				
OC	NCBI_TaxID=1613;				
OX	[1]				
RP	NUCLEOTIDE SEQUENCE [GENOMIC DNA].				
RA	STRAIN=CIP 102980T;				
RC	Kaminski P.A.;				
RT	"Characterization of the Lactobacillus fermentum N-deoxyribosyltransferase.";				
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.				
CC	FUNCTION: Catalyzes the cleavage of the glycosidic bond of 2'-deoxyribonucleosides and the transfer of the deoxyribosyl moiety to an acceptor purine or pyrimidine base [By similarity].				
CC	CATALYTIC ACTIVITY: 2-deoxy-D-ribose(1)-base(1) + base(2) = 2-deoxy-D-ribose(1)-base(2) + base(1).				
CC	- PATHWAY: Nucleotide metabolism.				
CC	- SIMILARITY: Belongs to the nucleoside deoxyribosyltransferase family.				
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.				
CC	EMBL; AY064168; AAL73115.1; -; Genomic DNA.				
DR	InterPro; IPR007710; N_deoxyrib_trans.				
DR	Pfam; PF05014; Nuc_deoxyrib_tr; 1.				
KW	Nucleotide metabolism; Transferase.				
FT	ACT SITE	103 103	Nucleophile (By similarity).		
SQ	SEQUENCE 168 AA; 18896 MW; 362162F43586C317 CRC64;				

Query March	24.6%	Score 209;	DB 1;	Length 168;
Best Local Similarity	36.4%;	Pred. No. 6.8e-10;		
Matches 55;	Conservative 23;	Mismatches 59;	Indels 14;	Gaps 7

QY	6	IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGRVDEHPEYLDHKV-W	64
Db	12	IYLATSPFNEQARIPQALQALENPVGVVHQ--PFDFQYKQARVDSDPAGVFGSLEW	69
QY	65	ATATYNDNLGIKTNIMLGVYIPEEDVDGLGMELGYALSOQK-YVLLVIPDEDYGK-PI	122
Db	70	QIATYNDNLNAVGTSDVCVALYDMQIDEGICMEIGMEFVALHKPIVLLPFTKKDKSAYEA	129
QY	123	NIM-----SWGVSNDNVIRKMSQLKDFNFKNP	147
Db	130	NIMLARGVTTW-LBPN--DFSLPKDFNFNHP	157

RESULT 7					
Q8RLY5	LACHE				
ID	Q8RLY5	LACHE	PRELIMINARY;	PRT;	167
AC	Q8RLY5;				

```
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Purine trans deoxyribosylase (Nucleoside deoxyribosyltransferase-I)
DE (EC 2.4.2.6).
DE Name=ptd; Synonyms=ndta;
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CRZ32; PubMed=11836245; DOI=10.1074/jbc.M11995200;
RX MEDLINE=21964041; PubMed=11836245; DOI=10.1074/jbc.M11995200;
RA Kaminski P.A.;
RT "Functional cloning, heterologous expression, and purification of two
RT different N-deoxyribosyltransferases from Lactobacillus helveticus.";
RL J. Biol. Chem. 277:14400-14407(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Okuyama K., Noguchi T.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY064166; AAL73113.1; -; Genomic DNA.
DR EMBL; AB076265; BAC00952.1; -; Genomic DNA.
DR PDB; 1S2D; X-ray; A/B/C=-.
DR PDB; 1S2G; X-ray; A/B/C=1-167.
DR PDB; 1S2I; X-ray; A/B/C=1-167.
DR PDB; 1S2L; X-ray; A/B/C=1-167.
DR PDB; 1S3F; X-ray; A/B/C=1-167.
DR GO; GO:0050144; F:transferase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR InterPro; IPR007710; N deoxyrib. trans.
DR Pfam; PF05014; Nuc deoxyrib tr; 1.
DR Glycosyltransferase; Transferase.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 167 AA; 18713 MW; 5540581511CB4787 CRC64;

Query Match 20.2%; Score 171.5; DB 2; Length 167;
Best Local Similarity 31.4%; Pred. No. 1.1e-06;
Matches 49; Conservative 34; Mismatches 58; Indels 15; Gaps 7;

QY 1 MPKTIYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDQYKGRVDEHPEY-- 58
DB 5 VPTGKIYLGFFYSDAQRAAKAKELLAKNPISL--AHVFFPDDGTD-PDEKNPBI 61
QY 59 LHKVWATATYNNDLNGIKTNDIMLGVIYIPDEEDVGLGMELGYALSQGYVLLV---IP 114
DB 62 IRSVWRDATYQNDLTGTSNATCGVFLYDMQDQDGSAPFSGFWRAMHKPVILVPFTEHP 121
QY 115 DEDYGKPINLM-SWGVS---DNVVKMSQLKDFNFK 146
DB 122 EKE--KKNNLMIAQGVTTIIDGNTEPEKLDYFNFE 155

RESULT 8
QSFNO_LACAC
ID QSFNO_LACAC PRELIMINARY; PRT; 167 AA.
AC QSFNO;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Putative deoxyribosyltransferase.
GN OrderedLocusNames=LBA1631;
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RX PubMed=15671160; DOI=10.1073/pnas.0409188102;
RA Altermann E., Russell W.M., Azcarate-Peril M.A., Barrangou R.;
RA Buck B.L., McAuliffe O., Southern N., Dobson A., Duong T., Callanan M.,
RA Lick S., Hamrick A., Cano R., Klaenhammer T.R.;
```

```
RT "Complete genome sequence of the probiotic lactic acid bacterium
RT Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912(2005).
DR EMBL; CP000033; AAV43444.1; -; Genomic DNA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR007710; N deoxyrib. trans.
DR Pfam; PF05014; Nuc deoxyrib tr; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 167 AA; 18933 MW; A80408F06C3E4D98 CRC64;

Query Match 18.2%; Score 154.5; DB 2; Length 167;
Best Local Similarity 32.5%; Pred. No. 3e-05;
Matches 49; Conservative 25; Mismatches 60; Indels 17; Gaps 7;

QY 6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDQYKGRVDEHPEY--LH 60
DB 10 IYLGTFYNDQDQARVYKARALLEQNTV--VRVHFFPQNF---VDEEKDPEADGLR 63
QY 61 DKVWATATYNNDLNGIKTNDIMLGVIYIPDEEDVGLGMELGYALSQGYVLLV--IPEDY 118
DB 64 SMTWRLATYNNDLSGIINATCGVFLYDMQDQDGSAPFSGFWRAMHKPVILVPFTNDPNK 123
QY 119 GKPINLM-SWGVS---DNVVKMSQLKDFNFK 145
DB 124 EKNNLMIAQGVTTIIDGNTELEKLTATYDFN 154

RESULT 9
Q9C173_LACLA
ID Q9C173_LACLA PRELIMINARY; PRT; 159 AA.
AC Q9C173;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein yefJb.
GN Names=yefJb; OrderedLocusNames=LL0493;
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471; DOI=10.1101/gr.GR-1697R;
RA Bolotin A., Wincker P., Mauder S., Jaillon O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL; AS006284; AAK04591.1; -; Genomic DNA.
DR PIR; E86686; E86686.
DR InterPro; IPR007710; N deoxyrib. trans.
DR Pfam; PF05014; Nuc deoxyrib tr; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 159 AA; 18295 MW; 98AE589D9C099C3 CRC64;
```

```
Query Match 16.1%; Score 136.5; DB 2; Length 159;
Best Local Similarity 26.7%; Pred. No. 0.00099;
Matches 43; Conservative 34; Mismatches 57; Indels 27; Gaps 7;

QY 6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDQYKGRVDEHPEYLDKVV-- 63
DB 11 VYLAAPFFSESQIKKVELLENALSKNTV--ANFFSPMRCQ-----HPESLPQVEA 60
QY 64 ----WATATYNNDLNGIKTNDIMLGVIYIPDEEDV--GLGMELGYALSQGYVLLVDPED 117
DB 61 FTPEWAKATMENDVNEVKNADIIVDFDQDTSGTAWELGYATALEKPTYLIRFEDT 120
QY 118 YGKPINLMSWGVSDN----VVKMSQLKDFNFKPRDFYEG 154
DB 121 I--PANIM--LTERNRAPFTQIEQVEEYDFLESKLIPYSG 156
```

RESULT 10

```

Q88SL4 LACPL
ID Q88SL4_LACPL PRELIMINARY; PRT; 157 AA.
AC Q88SL4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lp_3397.
GN OrderedLocustNames=lp_3397;
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566; DOI=10.1073/pnas.0337704100;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935262; CAD65499.1; -; Genomic_DNA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 157 AA; 17604 MW; 15E1FB38FE1F6746 CRC64;

Query Match 14.7%; Score 124.5; DB 2; Length 157;
Best Local Similarity 24.6%; Pred. No. 0.01; Mismatches 50; Indels 15; Gaps 4;
Matches 31; Conservative 30;

QY 6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKIRVDEHPEYLHDKVWA 65
DB 4 VYLAAPFDEAQRKIQKVSALLANPTINDGIFPEHQF----EERP-FGSRAMQ 56

QY 66 TATYNNDLNGIKTNDIMLGVI-----IPDEEDVGLGMEGLYALSQGYKIVLLVLPDEYDYG 119
DB 57 QYVYASDNQRHVADVVVAILDFOMTSATNSPDGTMFEIGAAVAETPFIIV--QFDAN 114

QY 120 KPINLM 125
DB 115 KELNLM 120

RESULT 11
Q04QAY0 LEIMA
ID Q04QAY0_LEIMA PRELIMINARY; PRT; 155 AA.
AC Q04QAY0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=LmjF23.1580;
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Friedlin;
RA Peacock C.S., Murphy L., Ivens A.C., Berriman M., Blackwell J.,
RA Smith D., Collins M., Foster N., Harris D., Oliver K., O'Neill S.,
RA Saunders D., Seeger K., Warren T., Apostolou Z., Bauser C., Beck A.,
RA Bianchetti G., Borzym K., Bothe G., Bruechi C., Ciaroni L.,
RA Dueterhoeft A., Fuchs M., Gabel C., Goffeau A., Hilbert H.,
RA Klages S., Kube M., Marra M., Masuy D., Mueller-Auer S., Pohl T.,
RA Purnelle B., Reinhardt R., Rieger M., Robben J., Schaefer M.,
RA Tosato V., Volckaert G., Wambutt R., Wedler H., Zimmermann M.,
RA Rajandream M., and Barrell B.G.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBSJ databases.
DR EMBL; CR05262; CAJ05145.1; -; Genomic_DNA.
KW Hypothetical protein.

SQ SEQUENCE 155 AA; 16887 MW; A9C05340F8D13F0 CRC64;

Query Match 12.5%; Score 106; DB 2; Length 155;
Best Local Similarity 27.6%; Pred. No. 0.39;
Matches 40; Conservative 16; Mismatches 47; Indels 42; Gaps 6;

QY 4 KTIYF-GAGWFTDRQNKAYKEAMEA-LKENPTIDLENSYVPLDNOYKGIKIRVDEHPEYLHDK 61
DB 5 KTIYIAGPAVHPDNGEAYNNHVRALLKEGVVPL-----IPVDNIATG----- 48

QY 62 KWMATATYNNDLNGIKTNDIMLGVIIP-----DEEDVGLGMEGLYALSQGYKIVLLVLPD--- 115
DB 49 ---ALSIRKNKIDMIRACDAVIADLSPEFSKPDGCTAFELGYAALGKLLLTETTTDTRP 105

QY 116 -----EDYKPINLM 125
DB 106 MVEYKGEGMAEGLSVENFGFLPFNLIM 130

RESULT 12
Q88X23 LACPL
ID Q88X23_LACPL PRELIMINARY; PRT; 146 AA.
AC Q88X23;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lp_1427.
GN OrderedLocustNames=lp_1427;
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566; DOI=10.1073/pnas.0337704100;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935256; CAD63892.1; -; Genomic_DNA.
DR InterPro; IPR007710; N_deoxyrib_trans.
DR Pfam; PF05014; Nuc_deoxyrib_tr; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 146 AA; 16608 MW; E33D94B52F7BE2A9 CRC64;

Query Match 11.3%; Score 96; DB 2; Length 146;
Best Local Similarity 25.3%; Pred. No. 2.6;
Matches 39; Conservative 28; Mismatches 63; Indels 24; Gaps 8;

QY 6 IYFGAGWFTDRQNKAYKEAMEALKENPTIDLENSYVPLDNOYKGIKIRVDEHPEYLHDKVWA 65
DB 4 IYIASPFSPSEQVTRVQLAEALQNPV--TDYFSPRLHQ-----DAQEQF-TKTWA 54

QY 66 TATYNNDLNGIKTNDIMLGVI-----YIPDEEDVGLGMEGLYALSQGYKIVLLVLPDEYDYGKPIN 123
DB 55 TEIFHRDWAQIAAAQAVIVTVLDFEAKNLDSTAYELGVATMRD---LPIIALQEKDEAVN 111

QY 124 LM-----SWGVSNDNVIRKMSQLKDFNFK-PRPDF 151
DB 112 LMITESLHWYTK-----QVSDPETYDFQQLPKGDF 141

RESULT 13
Q9EMR0 AMEPV
ID Q9EMR0_AMEPV PRELIMINARY; PRT; 1149 AA.
AC Q9EMR0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```

```

DE Name=AMV139;
GN Ansaeta moorei entomopoxvirus (AmEPV).
OS Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Betaentomopoxvirus.
OX NCBI_TaxID=28321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20396580; PubMed=10936094; DOI=10.1006/viro.2000.0449;
RA Bawden A.L., Glassberg K.J., Diggins J., Shaw R., Farmerie W.,
RA Moyer R.W.;
RT "Complete genomic sequence of the Ansaeta moorei entomopoxvirus:
RT analysis and comparison with other poxviruses.";
RL Virology 274:120-139(2000).
DR EMBL; AF250284; AAGO284.1; -; Genomic DNA.
SQ SEQUENCE 1149 AA; 135479 MW; 7E4EE54A13F87691 CRC64;

Query Match 11.3%; Score 96; DB 2; Length 1149;
Best Local Similarity 23.7%; Pred. No. 27;
Matches 42; Conservative 29; Mismatches 62; Indels 44; Gaps 10;

QY 13 FTD--RONKAYKEAMEALKENPTIDLEN-----SYVPLD-----NOYKGIKRVDEH 55
DB 12 FTFIIRNLPRNLDSKKNIIINNETYKIEELFKYIYTHPLDLLTIRDISNADKDEY 71

QY 56 PEYLHDKVWATATYNN-DL--NGIKTNDIMLGV-----YIPDEEDVGLGMEGLYALSQ-- 105
DB 72 VKQPVNNLYLRAYNEMDFIKNIRYDDKVSIIINEINYPPEHTSBELKYRLSHYESER 131

QY 106 ---GKYVLLV-IPDEDYG-----KPINLMSWGSVDNVKMSQLKDFNFKPRDFY 152
DB 132 IRGRGVVTFSGVPDNGYGYLLSQSDPSKVIWAIVDNYLMID-----NEDKDFDY 181

RESULT 14
Q5ZMQ1.CHICK PRELIMINARY; PRT; 1029 AA.
AC Q5ZMQ1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Hypothetical protein.
GN ORFNames=RCUMB04_1h14;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kuter S., Blagodatski A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursa lymphocytes to facilitate
RT gene function analysis.";
RL Genome Biol. 6:R6-R6(2005).
DR EMBL; AJ719333; CAG30992.1; -; mRNA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR012961; DSHCT.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF08148; DSHCT; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW Hypothetical protein.
SQ SEQUENCE 1029 AA; 116048 MW; EC8909BE178AF562 CRC64;

```

```

Query Match 11.2%; Score 95; DB 2; Length 1029;
Best Local Similarity 22.4%; Pred. No. 29;
Matches 45; Conservative 32; Mismatches 56; Indels 68; Gaps 9;

QY 1 MPKKTIVF-----GAGWFTDRQNKAYKEAMEAL-----KENPTID-- 35
DB 478 MPARTVLTFSASKFDGKDFRWISSGEYIQMSGAGRRGMDDRGIVILMVDERKMSPTIGKQ 537

QY 36 -LENSYVPLDNQY-----KGIRVDE-HPEYLHDKVWATATYNNDLNGI-----KTND 80
DB 538 LLKGSADPLNSAFHLTYNNVNLRLRVEEINPEYMLEKSFYQFOHYRAIPGVVEKVNKLEE 597

QY 81 IMLGVVPIPEEDV-----GLGMEGLYALSOGKYVLLVI-----PDEDYG 119
DB 598 QYNKIVIPNEENVVIYKIRQRLAKLGKEIEYIHKPKYCLPFLQPLQRLVKVKNEDDDFG 657

QY 120 KPINLMSWGSVDNVKMSQLK 140
DB 658 -----WGVVNVNFSKSNVK 671

RESULT 15
Q4ZQ06_PLABE PRELIMINARY; PRT; 264 AA.
AC Q4ZQ06;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein (Fragment).
GN ORFNames=PB000574.01.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAI01001273; CAH96134.1; -; Genomic DNA.
KW Hypothetical protein.
FT NON TER 264 264
SQ SEQUENCE 264 AA; 31399 MW; EDD7A47344723240 CRC64;

Query Match 11.0%; Score 93.5; DB 2; Length 264;
Best Local Similarity 27.1%; Pred. No. 8.3;
Matches 39; Conservative 24; Mismatches 52; Indels 29; Gaps 7;

QY 19 KAYKEAMEALKENPTIDLENSYVPLDNQYKIGR-VDEHPEYLHDKVWATATYNNDLNGIK 77
DB 34 RSYLQAMECLKNNEM---NTYKKNKQKYLDTKNLEDSDDLHKIKWINNNSINNKNSFK 90

QY 78 T-----NDIMLGVIIPDEEDVGLGMEGLYALSOGKYVLLVIPDEDYCKPINLMSWGSVDN 132
DB 91 TIQKYGNNI--GFLKDER-----RLNVALTRAKDSLWIIGDK-----TNLQKNSTWDS 136

QY 133 VIKMSQLK-----DFNFKPRFD 150
DB 137 LIKNATARNCYVNLNLFDRSTKD 160

Search completed: December 23, 2005, 22:04:15
Job time : 238 secs

```